```
OM protein - protein search, using sw model

            GenCore version 5.1.6
            Copyright (c) 1993 - 2005 Compugen Ltd.

Run on:    February 26, 2005, 18:36:46 ; Search time 176 Seconds
                     (without alignments)
                     2417.828 Million cell updates/sec

Title:     US-10-701-038-2
Perfect score: 4310
Sequence: 1 PSGDIQDSQQDEQQOQQ.........DRGADINLDGDHNILFDQCL 831

Scoring table: BLOSUM62
           Gapop 10.0 , Gapext 0.5

Searched:  1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database:  UniProt 03:*
           1: uniprot_sprot:*
           2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4310	100.0	1421	2	Q9APN7	Q9apn7 ehrlichia c
2	442	10.3	1179	2	Q8AX92	Q8ax92 fugu rubrip
3	421	9.8	1059	1	AN28_HUMAN	O15084 homo sapien
4	390.5	9.1	1719	2	Q13768	Q13768 homo sapien
5	390.5	9.1	1856	2	Q99407	Q99407 homo sapien
6	390.5	9.1	1880	1	ANK1_HUMAN	P16157 homo sapien
7	387	9.0	1913	2	Q9GRV5	Q9grv5 caenorhabdi
8	382.5	8.9	1848	2	Q61302	Q61302 mus musculu
9	381.5	8.9	1862	1	ANK1_MOUSE	Q02357 mus musculu
10	378.5	8.8	1943	2	Q61307	Q61307 mus-musculu
11	377.5	8.8	4377	1	ANK3_HUMAN	Q12955 homo sapien
12	377	8.7	1136	2	Q9N180	Q9n180 bos taurus
13	373.5	8.7	687	2	Q6ZT57	Q6zt57 homo sapien
14	371.5	8.6	1726	2	Q8VC68	Q8vc68 mus musculu
15	371	8.6	1619	2	Q9VMR4	Q9vmr4 drosophila
16	371	8.6	1712	2	Q7KTN8	Q7ktn8 drosophila
17	370.5	8.6	1887	2	Q7Z3G4	Q7z3g4 homo sapien
18	369.5	8.6	1863	2	Q7Z3L5	Q7z3l5 homo sapien
19	368.5	8.5	2622	2	O70511	O70511 rattus norv
20	368.5	8.5	3924	1	ANK2_HUMAN	Q01484 homo sapien
21	365.5	8.5	692	2	Q6NRI3	Q6nri3 xenopus lae
22	364.5	8.5	919	2	Q8N8A2	Q8n8a2 homo sapien
23	364	8.4	933	2	Q73J83	Q73j83 treponema d
24	358.5	8.3	1219	2	Q8C8R3	Q8c8r3 mus musculu
25	358	8.3	1614	2	Q7T1G6	Q7t1g6 brachydanio
26	355.5	8.2	843	2	P97582	P97582 rattus norv
27	354.5	8.2	531	2	Q8C036	Q8c036 mus musculu
28	354	8.2	1076	2	Q8BTI7	Q8bti7 mus musculu
29	353.5	8.2	2206	2	Q7QAV8	Q7qav8 anopheles g
30	350.5	8.1	1441	2	Q7Q8L0	Q7q8l0 anopheles g
31	350	8.1	1088	2	Q13484	Q13484 homo sapien
32	345.5	8.0	1762	2	O88521	O88521 rattus norv
33	345	8.0	542	2	Q73I01	Q73i01 wolbachia p
34	345	8.0	791	2	Q9UFA4	Q9ufa4 homo sapien
35	344	8.0	1401	1	LATA_LATMA	P23631 latrodectus
36	342	7.9	934	2	O83807	O83807 treponema p
37	340	7.9	1145	2	Q7PE28	Q7pe28 anopheles g
38	340	7.9	1501	2	Q7QKD3	Q7qkd3 anopheles g
39	340	7.9	1549	2	Q9V4B1	Q9v4b1 drosophila
40	339.5	7.9	1549	2	Q24241	Q24241 drosophila
41	338.5	7.9	745	2	Q8B225	Q8b225 m mus muscu
42	338	7.8	1098	2	Q61304	Q61304 mus musculu
43	336	7.8	579	2	Q8IZ72	Q8iz72 homo sapien
44	336	7.8	1159	2	Q9NCP8	Q9ncp8 drosophila
45	336	7.8	1571	2	Q7KU92	Q7ku92 drosophila

ALIGNMENTS

```
RESULT 1
Q9APN7
ID   Q9APN7             PRELIMINARY;       PRT;   1421 AA.
AC   Q9APN7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   200 kDa immunoreactive glycoprotein.
OS   Ehrlichia canis.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Anaplasmataceae; Ehrlichia.
OX   NCBI_TaxID=944;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20579049; PubMed=11136790;
RA   McBride J.W., Corstvet R.E., Breitschwerdt E.B., Walker D.H.;
RT   "Immunodiagnosis of Ehrlichia canis infection with recombinant
RT   proteins.";
RL   J. Clin. Microbiol. 39:315-322(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   McBride J.W., Comer J.E., Walker D.H.;
RT   "Novel Immunoreactive Glycoprotein Orthologs of Ehrlichia spp.";
RL   Ann. N. Y. Acad. Sci. 0:0-0(2003).
DR   EMBL; AF252298; AAK01145.2; -.
DR   HSSP; O75832; 1QYM.
DR   InterPro; IPR002110; ANK.
DR   Pfam; PF00023; Ank; 20.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 21.
DR   PROSITE; PS50088; ANK_REPEAT; 8.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
SQ   SEQUENCE 1421 AA; 152990 MW; 6063E9057A375D40 CRC64;

Query Match        100.0%;  Score 4310;  DB 2;  Length 1421;
Best Local Similarity 100.0%;  Pred. No. 1.1e-228;
Matches 831;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    1 PSGDIQDSQQDEQQOQQGAVGNSPIERERVAAPESEDLYTVIIPKGKRTAAPI 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16 PSGDIQDSQQDEQQOQQGAVGNSPIERERVAAPESEDLYTVIIPKGKRTAAPI 75

Qy   61 LERKSPTTEPKVEDDEDLPPTLPPRTFSGEGYDDVGVSMPTVSRGIYQPPIVQDSNLYSS 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 LERKSPTTEPKVEDDEDLPPTLPPRTFSGEGYDDVGVSMPTVSRGIYQPPIVQDSNLYSS 135

Qy  121 IGGVPQEAQYDAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDVRNAVLGNKEIKEE 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136 IGGVPQEAQYDAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDVRNAVLGNKEIKEE 195

Qy  181 WLTTSGPVRDIADRIVASKGDLSEDQVEILDIIFMNESEIAEGISNPLHADVDNNPVKG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  196 WLTTSGPVRDIADRIVASKGDLSEDQVEILDIIFMNESEIAEGISNPLHADVDNNPVKG 255
```

```
Qy    71  KVEDDEDLPPTLPPRTFSGEGYDDVGV--------SMPTVSRGIYQPPIVQDSNLYSSIG  122
              :|:||                            :|::|
Db   279  EVEDEPSPSPPPQTKSSPERRSHPVCRKAATHQDKPSLLRAIFNVD------------    326

Qy   123  GVPQEAQYDAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDVRNAVLGNKEIKE-    179
              :|
Db   327  --PEEV----------RSLILK----KEDVNVQDNEKRTPLHAA--AYLGDABIIELLI   367

Qy   180  ----------------EWLTTSGPVRDIADRIVASKGDiSEDQVEEILDIIFMNESEIAEGISN-  227
                            ||:   :::    ||      : :::
Db   368  LSGARVNAKDNKWLT---PLH----RAVAS----CSEDAVA------VLLKHSADVNARDKNW   413

Qy   228  --PLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKAL------GEVENDEGDLGANAY   279
              :|:                      |:|:     :    :|::|:|:||
Db   414  QTPLHVAASNKAVRCAEALVPLLSNVNVSDRAGR--TALHHAAFSGHVEVNE----VNAY   467

Qy   280  NVLDSEGNLPLHHAAKN----CTGDKLK------------------------LCMEKT   309
              |:|  ||||:|:                                  |:|
Db   468  ------GNTPLHLACYNGQDVVVGELIQAGAKVNQENERGFSPLHFASSSRQGALCQELL   521

Qy   310  KTD--FIDTANFANQSPLHIITQ--KPDCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVE   365
              :|   |:||:|||||||:        |:|          ::
Db   522  LTHGAHINMQSKDNRTPLHMAATHGRFSCCS----QALIQNGADIDCEDKNRNTALHIAAR   577

Qy   366  HLPPVILKGVMDHVKNS---------------------SEFQDL------VNDPYFGNTI   399
              |:  ::      :::                      ::
Db   578  QGHELIITALIKHGANSARRGVHGMFPLHLAALSGFSDCCRKLLSSGFDTDPEIGRTC      637

Qy   400  AHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCG-----  454
              ::  ||:    ||:|:|:|:||:|::    ||
Db   638  LHAAAAGGN--LECLNLLLKIGADFNRKDNFGRTPLHYASANCNYQCVFALVGSGASVNV  695

Qy   455  ----------------------------------------------------------  454

Db   696  LDQRGCNPLHYAAAADTEGKGHQDCVALLLHHGASPMTRDYAHKKTAIHAAAMNGHQECL  755

Qy   455  IDVNSQDVNGDTPLHLAVEGGSMETVLAVLN--------------             485
              :|::|:|:|:||:||::|:||:::::|
Db   756  RLLMSHSQHLDVDAQDINGQTPLMLAVLNGHTECVYSLLSQGASVENQDRWGRTALHRGV  815

Qy   486  ------ORGADVSVQNNDCVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKS        533

Db   816  VTGQEECVEALLQRGANVCVKDIQGRSPLHLASACGRVGALGALLQASSTSHAH-----  869

Qy   534  LLMEDYKGFTPLHFVAGGSGRDTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFE  593
              ::|         ||:||   :
Db   870  --LTDNQGYTPLHWACYNGT----------GYDSCVEVLLDQEVFKQ---------     904

Qy   594  SENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDS-LHLPSERGYNA  652
              ::                :|:||
Db   905  ---IKGN----------SFSPLHCAVINDNEGVAEMLIESMGTNIINTSDSKGRTP      947

Qy   653  MQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGRGDNEVVRGLVQHGIDIN  712
              :::|:     |
Db   948  LHAAAFSDHVECVSLLLSHGAQANV-ADTRLCRTPLMAALNGQNTVEVLVNSAKVDLT     1006

Qy   713  QRMGSDKNTVLHYAISKGDSFLVQKILAHTGVD--VNCENNLGQTPLHLAVEGGDPKIVS  770
              :     :||:|||| |:|: |::||||:|| |    || ||:|||||||:||:|||:|:
Db  1007  LQ-DAHRNTALHLACSKGHETCALLILEKIGDRNLINCTNAALQTPLHVAARRGLTVVVQ  1065

Qy   771  SLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIV                         807
              ::::  |:::  :::    |
Db  1066  ELLGKEASVLAVDENGYTPALSC-TPNRDVADCLALI                       1101
```

```
Qy   241  AKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGD  300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   256  AKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGD  315

Qy   301  KLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPL  360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   316  KLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPL  375

Qy   361  HHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPYFGNTIAHYAVKNKNADLTLFNMLKAS   420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   376  HHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPYFGNTIAHYAVKNKNADLTLFNMLKAS   435

Qy   421  GADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETV  480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   436  GADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETV  495

Qy   481  LAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKSLLMEDYK  540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   496  LAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKSLLMEDYK  555

Qy   541  GFTPLHFVAGGSGRDTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGS  600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   556  GFTPLHFVAGGSGRDTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGS  615

Qy   601  PNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFG  660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   616  PNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFG  675

Qy   661  DKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGRGDNEVVRGLVQHGIDINQRMGSDKN  720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   676  DKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGRGDNEVVRGLVQHGIDINQRMGSDKN  735

Qy   721  TVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVN  780
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   736  TVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVN  795

Qy   781  RLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADINLDGDHNILFDQCL          831
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   796  RLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADINLDGDHNILFDQCL          846
```

RESULT 2
Q8AX92
Q8AX92;
ID   Q8AX92   PRELIMINARY;   PRT;   1179 AA.
AC   Q8AX92;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetraodontoidea; Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kosan C., Kunz J.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF527754; AAQ15006.1; -.
DR   HSSP; P55271; 1D9S.
DR   InterPro; IPR002110; ANK.
DR   Pfam; PF00023; Ank; 23.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 22.
DR   PROSITE; PS50088; ANK_REPEAT; 16.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   ANK repeat; Hypothetical protein.
SQ   SEQUENCE 1179 AA; 127675 MW; ABFF2F2B2CC14EF2 CRC64;

Query Match   10.3%; Score 442; DB 2; Length 1179;
Best Local Similarity 22.5%; Pred. No. 1.2e-15;
Matches 211; Conservative 102; Mismatches 310; Indels 314; Gaps 34;

RESULT 3
AN28_HUMAN
ID   AN28_HUMAN   STANDARD;   PRT;   1059 AA.
AC   O15084;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)

```
Db   55 AVLGDAEIIELLILSGARVNAKDSKWLT---PLH----RAVASCS-------EEAVQVLL 100
             |: |||||         |
Qy  216 MNESEIAEGISN--PLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEG 272

Db  101 KHSADVNARDKNWQTPLHIAAANKAVKCAEALVPLLSNV-------------------- 139

Qy  273 DLGANAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKP 332

Db  140 ------NVSDRAGRTALHHAAFSGHGEMVKLLLSRGAN--INAFDKKDRRAIHWAAYMG 190

Qy  333 DCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDP 392

Db  191 HIEVVKL--LVSHGAEVTCKDKKSYTPLHAAAS-------SGMISVVKYLLDLGVDMNEP 241

Qy  393 DYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIH-VASSNGKANAVSGLV 451

Db  242 NAYGNTPLHVACYN-GQDVVV-NELIDCGAIVNQKNEKGFTPLHFAAASTHGALCLELLV 299

Qy  452 SCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDI 511

Db  300 GNGADVNMKSKDGKTPLHMTALHGRFSRSQTII-QSGAVIDCEDKNGNTPLHIAARYGHE 358

Qy  512 GVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLA 571

Db  359 LLINTLITS-----GADTAKRGI-----HGMFPLHLAALSGFSD---CCRKLLSSGFDID 405

Qy  572 TI----RAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSN 627

Db  406 TPDDFGRTCLHAAAAGGNL-------ECLNLLNTGADF---NKKDKFGRSPLHYAAANCN 456

Qy  628 HNVMKDIIN--FVGD--------SLH---------LPSERGYNAMQ 654

Db  457 YQCLFALVGSGASVNDLLDERGCTPLHYAATSDTDGKCLEYLLRNDANPGIRDKQGYNAVH 516

Qy  655 VAALFGDKEAVKMLAKSAKPSDLNFKTS-----------ATPTPLNLACLRGDNEVVRGL 703

Db  517 YSAAYGHRLCLQLIA-SETPLDVLMETSGTDMLSDSDNRATISPLHLAAYHGHHQALEVL 575

Qy  704 VGQHGIDINQRMGSD------------------KNTVLHYAISKG 730

Db  576 V-QSLLDLDVRNSSGRTPLDLAAFKGHVECVDVLINQGASILVKDYILKRTPIHAAATNG 634

Qy  731 DSFLVQKILAH----TGVDVNCENNLGQTPLHIAVEGGDPKIVSSLLKAGAVVNRLDDNG 786

Db  635 HSECLRLLIGNAEPQNAVDIQDGN--GQTPLMLSVLNGHTDCVYSLLNKGANVDAKDKWG 692

Qy  787 RSVLSSAIVPGRKEKGVLGIVNKLLDRGA 815

Db  693 RTALHRGAVTGHEE-------CVDALLQHGA 716
```

RESULT 4
Q13768
ID   Q13768   PRELIMINARY;   PRT;   1719 AA.
AC   Q13768;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Alt. ankyrin (Variant 2.2).
OS   Homo sapiens (Human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Hematopoietic;
RX   MEDLINE=90158830; PubMed=2137557;
RA   Lux S.E., John K.M., Bennett V.;
RT   "Analysis of cDNA for human erythrocyte ankyrin indicates a repeated
RT   structure with homology to tissue-differentiation and cell-cycle
RT   control proteins.";
RL   Nature 344:36-42(1990).
DR   EMBL; X16609; CAA34611.1; -.

DE   Ankyrin repeat domain protein 28 (Fragment).
GN   Name=ANKRD28; Synonyms=KIAA0379;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97349984; PubMed=9205841;
RA   Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
CC   -!- SIMILARITY: Contains at least 27 ANK repeats.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AB002377; BAA20833.2; -.
DR   HSSP; P80144; 2MYO.
DR   IntAct; O15084; -.
DR   Genew; HGNC:29024; ANKRD28.
DR   InterPro; IPR002110; ANK.
DR   Pfam; PF00023; Ank; 28.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 28.
FT   PROSITE; PS50297; ANK_REP_REGION; 1.
FT   PROSITE; PS50088; ANK_REPEAT; 24.
KW   ANK repeat; Repeat.
FT   NON_TER   1   1
FT   REPEAT   46   75   ANK 1.
FT   REPEAT   79   108   ANK 2.
FT   REPEAT   112   141   ANK 3.
FT   REPEAT   145   174   ANK 4.
FT   REPEAT   178   207   ANK 5.
FT   REPEAT   211   240   ANK 6.
FT   REPEAT   244   273   ANK 7.
FT   REPEAT   277   307   ANK 8.
FT   REPEAT   311   340   ANK 9.
FT   REPEAT   344   373   ANK 10.
FT   REPEAT   377   406   ANK 11.
FT   REPEAT   410   439   ANK 12.
FT   REPEAT   443   472   ANK 13.
FT   REPEAT   476   506   ANK 14.
FT   REPEAT   510   540   ANK 15.
FT   REPEAT   544   584   ANK 16.
FT   REPEAT   588   617   ANK 17.
FT   REPEAT   622   651   ANK 18.
FT   REPEAT   658   687   ANK 19.
FT   REPEAT   691   720   ANK 20.
FT   REPEAT   724   753   ANK 21.
FT   REPEAT   761   790   ANK 22.
FT   REPEAT   793   823   ANK 23.
FT   REPEAT   828   857   ANK 24.
FT   REPEAT   861   891   ANK 25.
FT   REPEAT   895   924   ANK 26.
FT   REPEAT   931   960   ANK 27.
SQ   SEQUENCE   1059 AA;   113465 MW;   C1F55B6CFB494770 CRC64;

```
Query Match             9.8%;    Score 421;  DB 1;  Length 1059;
Best Local Similarity  25.2%;    Pred. No. 1.5e-14;
Matches  189;  Conservative 101;  Mismatches 269;  Indels 190;  Gaps 29;

Qy  170 AVLGNKEIKE------------EWLITTSGPVRDIADRIVASKGDLSEDQVEEILDIIF 215
```

```
Qy    821 GDHNLFPDQ 829
              :
Db    800 DKHRMSFPE 808
```

RESULT 5
Q99407

```
ID   Q99407          PRELIMINARY;      PRT;   1856 AA.
AC   Q99407;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Ankyrin.
GN   Name=ANK1;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97382244; PubMed=9235914; DOI=10.1074/jbc.272.31.19220;
RA   Gallagher P.G., Tse W.T., Scarpa A.L., Lux S.E., Forget B.G.;
RT   "Structure and organization of the human ankyrin-1 gene. Basis for
RT   complexity of pre-mRNA processing.";
RL   J. Biol. Chem. 272:19220-19228(1997).
DR   EMBL; U50092; AAB47805.1; JOINED.
DR   EMBL; U50093; AAB47805.1; JOINED.
DR   EMBL; U50094; AAB47805.1; JOINED.
DR   EMBL; U50095; AAB47805.1; JOINED.
DR   EMBL; U50096; AAB47805.1; JOINED.
DR   EMBL; U50097; AAB47805.1; JOINED.
DR   EMBL; U50098; AAB47805.1; JOINED.
DR   EMBL; U50099; AAB47805.1; JOINED.
DR   EMBL; U50133; AAB47805.1; .
DR   EMBL; U50100; AAB47805.1; JOINED.
DR   EMBL; U50102; AAB47805.1; JOINED.
DR   EMBL; U50103; AAB47805.1; JOINED.
DR   EMBL; U50104; AAB47805.1; JOINED.
DR   EMBL; U50105; AAB47805.1; JOINED.
DR   EMBL; U50106; AAB47805.1; JOINED.
DR   EMBL; U50107; AAB47805.1; JOINED.
DR   EMBL; U50108; AAB47805.1; JOINED.
DR   EMBL; U50110; AAB47805.1; JOINED.
DR   EMBL; U50111; AAB47805.1; JOINED.
DR   EMBL; U50112; AAB47805.1; JOINED.
DR   EMBL; U50113; AAB47805.1; JOINED.
DR   EMBL; U50114; AAB47805.1; JOINED.
DR   EMBL; U50115; AAB47805.1; JOINED.
DR   EMBL; U50116; AAB47805.1; JOINED.
DR   EMBL; U50117; AAB47805.1; JOINED.
DR   EMBL; U50118; AAB47805.1; JOINED.
DR   EMBL; U50119; AAB47805.1; JOINED.
DR   EMBL; U50120; AAB47805.1; JOINED.
DR   EMBL; U50121; AAB47805.1; JOINED.
DR   EMBL; U50122; AAB47805.1; JOINED.
DR   EMBL; U50123; AAB47805.1; JOINED.
DR   EMBL; U50124; AAB47805.1; JOINED.
DR   EMBL; U50125; AAB47805.1; JOINED.
DR   EMBL; U50126; AAB47805.1; JOINED.
DR   EMBL; U50127; AAB47805.1; JOINED.
DR   EMBL; U50128; AAB47805.1; JOINED.
DR   EMBL; U50129; AAB47805.1; JOINED.
DR   EMBL; U50130; AAB47805.1; JOINED.
DR   EMBL; U50131; AAB47805.1; JOINED.
DR   EMBL; U50132; AAB47805.1; JOINED.
DR   HSSP; P16157; 1N11.
DR   GO; GO:0016323; C:basolateral plasma membrane; ISS.
DR   GO; GO:0005856; C:cytoskeleton; NAS.
DR   GO; GO:0008093; F:cytoskeletal adaptor activity; ISS.
DR   GO; GO:0019899; F:enzyme binding; ISS.
DR   GO; GO:0030507; F:spectrin binding; ISS.
DR   GO; GO:0005200; F:structural constituent of cytoskeleton; ISS.
DR   GO; GO:0006887; P:exocytosis; NAS.
DR   GO; GO:0045199; P:maintenance of epithelial cell polarity; ISS.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000488; Death.
DR   InterPro; IPR011029; DEATH_like.
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00023; Ank; 23.
DR   Pfam; PF00531; Death; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 23.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 20.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
KW   ANK repeat.
SQ   SEQUENCE   1719 AA;  189010 MW;  F63465D16D975CBF CRC64;
```

```
Query Match     9.1%;
Best Local Similarity   26.0%;
Matches 174;  Conservative 100;  Mismatches 246;  Indels 149;  Gaps 31;

Score 390.5;  DB 2;  Length 1719;
Pred. No. 1.5e-12;

Qy  194 RIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVK------GA---- 241
            |::.:||..|| .:: :  |...:   ...::
Db  256 RLLLDRGAQIETKTKDEL-----------------TPLHCAARNGHVRISEILLDHGAPIQA 300

Qy  242 --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT 298
          |:  || || |               |              :
Db  301 KTKNGLSPIHMAAAQGDHLD--CVRLLQYDAEIDDI------TLD---HLTPLHVAAHCG 349

Qy  299 GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDFGDGK 357
        |:|| |:|    |    |||    :||||     : |::  | | :
Db  350 HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL 405

Qy  358 NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT-IAHYAVKNK 407
         |||.|  :  |        :|   :      : |          |:: |:: |  :|:|
Db  406 TPLHVASFMGHLPVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK 460

Qy  408 NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP 467
                         |  :|   :  |  ||::|  .|||:  : ::|:|:||:  |:|
Db  461 ------AKVNAKAKDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP 506

Qy  468 LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK 525
        |||| .|| .|||||.| |:.|   :  : | ||.::|||||: ||  ||  |.| |||:
Db  507 LHIAAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA 565

Qy  526 GEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE 585
        |.:     :|:||||:| ||| |                :|
Db  566 GKN--------GLTPLHVAVHHNNLDIVKL---- 587

Qy  586 LVNLGDFESENILGSPNAKPLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP 645
        |:. |:.|:  :.|.     .:|  |::    |  :| :||:||:::     ||||:| :
Db  588 LLPRG------GSP------HSPAWN-GYTPLHIAAKQNQVEVARSLLQY-GGSANAE 631

Qy  646 SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG 705
        |.:|  .:||| :||::| |::|      :  .  .| : |||:.| |:|:| ||  |:|
Db  632 SVQGVTPLHLAAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHLVAQEGHVPVADVLI- 688

Qy  706 QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG 763
        :||   |:|      :||||:|:| :|:  |::| ||:: .||||   :| .||||:| :
Db  689 KHGVMVDATTRMG--YTPLHVASHYGNIKLIVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ 744

Qy  764 GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL---LDRGADINLD 820
        |  .||: ||| ||: |  .   . ||:::|    :::||: :  .:    :|.::|:||
Db  745 GHTDIVTLLLKNGASPNEVSSDGTTPLAIA------KRLGVISVTDVLKVVTDETSFVLVS 799
```

```
DR   GO; GO:0005200; F:structural constituent of cytoskeleton; ISS.
DR   GO; GO:0045199; P:maintenance of epithelial cell polarity; ISS.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000488; Death.
DR   InterPro; IPR011029; DEATH_like.
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00023; Ank; 22.
DR   Pfam; PF00531; Death; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 22.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 20.
DR   PROSITE; PS50297; ANK REP REGION; 1.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
KW   ANK repeat.
SQ   SEQUENCE   1856 AA;   203465 MW;   4A4E3A794DD6401F CRC64;
```

```
Query Match             9.1%;  Score 390.5;  DB 2;  Length 1856;
Best Local Similarity  26.0%;  Pred. No. 1.6e-12;
Matches  174;  Conservative 100;  Mismatches 246;  Indels 149;  Gaps  31;

Qy    194  RIVASKGDLSEDQVEBILDIIFMNESEIAEGISNPLHADVDNNPVK---------GA----- 241
                :|:::||                                           :|
Db    223  RLLLDRGAQIETKTKDEL------------TPLHCAARNGHVRISEILLDHGAPIQA 267

Qy    242  --KNVMTLMHLJVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNPLPLHHAAKNCT 298
             ||:|:: ||   :|:| : |:: ||:|    :|     ||  : :|  ||
Db    268  KTKNGLSPIHMAAQGDHLD--CVRLLLQYDAEIDDI------TLD---HLTPLHVAAHCG 316

Qy    299  GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK 357
           ||:|  ||:   : | :|    :::::|||||||: :: ||:::|||::||||||||||
Db    317  HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL 372

Qy    358  NPLHHA--VEHLPPV---ILKGVDHVKNSSEFQDLVNDPDYF---GNT-IAHYAVKNK 407
             ||||    |:|||     : |::| : ::| :::| : :|    :|: || ||||
Db    373  TPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK 427

Qy    408  NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP 467
                   :|:| : :| :|| | ||| ||:|||:: :
Db    428  ---------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTRAGHTP 473

Qy    468  LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK 525
           |||| :|| :| |: | |  ||  : ||| ||| |:||: | | ||| |   |:|||:
Db    474  LHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA 532

Qy    526  GEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE 585
              ::                ||||:||||
Db    533  GKN--------------GLTPLHVAVHHNNLDIVKL-------------------------- 554

Qy    586  LVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNPVGDSLHLP 645
              ::| ||: :|  |                   |: ||:  ::|:  :  :
Db    555  LLPRG-------GSP------HSPAWN-GYTPLHIAAKQNVEVARSLLQY-GGSANAE 598

Qy    646  SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG 705
           | || ||| || ||||  :   ||:|   :: : :| || ||||:||:|||::|| |
Db    599  SVQGVTPLHLAAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHLVAQEGHVPVADVLI-- 655

Qy    706  QHG--IDINQRMGSDKNTVLHYAISKGDSPLVQKILAHTGVDVNCENNLGQTPLHLAVEG 763
              ::  ::||::  ::||:|||||||:|||:||::|: :||:|||:|:::||||||||
Db    656  KHGVMVDATTRMG--YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ 711

Qy    764  GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL--LDRGADINLD 820
           || |:|||:|||||:|||| ||||:||:|:||||:|:|||:||:|||   |||||||:||
Db    712  GHTDIVTLLLKNGASPNEVSSDGTTPLAIA-----KRLGYISVTDVLKVVTDETSPVLVS 766

Qy    821  GDHNILFDQ 829
           |||:: ||
Db    767  DKHRMSFPE 775
```

RESULT 6

```
ANK1_HUMAN
ID   ANK1_HUMAN             STANDARD;      PRT;   1880 AA.
AC   P16157;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Ankyrin 1 (Erythrocyte ankyrin) (Ankyrin R).
GN   Name=ANK1; Synonyms=ANK;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.; PARTIAL SEQUENCE, AND VARIANTS.
RC   TISSUE=Hematopoietic;
RX   MEDLINE=90158830; PubMed=2137557;
RA   Lux S.E., John K.M., Bennett V.;
RT   "Analysis of cDNA for human erythrocyte ankyrin indicates a repeated
RT   structure with homology to tissue-differentiation and cell-cycle
RT   control proteins.";
RL   Nature 344:36-42(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90175370; PubMed=1689849;
RA   Lambert S., Yu H., Prchal J.T., Lawler J., Ruff P., Speicher D.,
RA   Cheung M.C., Kan Y.W., Palek J.;
RT   "cDNA sequence for human erythrocyte ankyrin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:1730-1734(1990).
RN   [3]
RP   VARIANT HS ILE-462.
RX   MEDLINE=96225450; PubMed=8640229;
RA   Eber S.W., Gonzalez J.M., Lux M.L., Scarpa A.L., Tse W.T.,
RA   Dornwell M., Herbers J., Kugler W., Oezcan R., Pekrun A.,
RA   Gallagher P.G., Schroeter W., Forget B.G., Lux S.E.;
RT   "Ankyrin-1 mutations are a major cause of dominant and recessive
RT   hereditary spherocytosis.";
RL   Nat. Genet. 13:214-218(1996).
CC   -!- FUNCTION: Attach integral membrane proteins to cytoskeletal
CC       elements; bind to the erythrocyte membrane protein band 4.2, to
CC       Na-K ATPase, to the lymphocyte membrane protein GP85, and to the
CC       cytoskeletal proteins fodrin, tubulin, vimentin and desmin.
CC       Erythrocyte ankyrins also link spectrin (beta chain) to the
CC       cytoplasmic domain of the erythrocytes anion exchange protein;
CC       they retain most or all of these binding functions.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic surface of erythrocytic plasma
CC       membrane.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1; Synonyms=2.1;
CC         IsoId=P16157-1; Sequence=Displayed;
CC       Name=2; Synonyms=2.2;
CC         IsoId=P16157-2; Sequence=VSP_000264, VSP_000265;
CC       Name=3;
CC         IsoId=P16157-3; Sequence=VSP_000266;
CC   -!- PTM: Regulated by phosphorylation.
CC   -!- PTM: Palmitoylated.
CC   -!- DISEASE: Defects in ANK1 are a cause of hereditary spherocytosis
CC       (HS) [MIM:182900]. Inheritance can be autosomal dominant or
CC       recessive.
CC   -!- SIMILARITY: Contains 23 ANK repeats.
CC   -!- SIMILARITY: Contains 1 death domain.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X16609; CAA34610.1; -.
DR   EMBL; M28880; AAA51732.1; -.
```

```
DR   PIR; A35049; A35049.
DR   PIR; S08275; SJHUK.
DR   PDB; 1N11; X-ray; A=390-826.
DR   Genew; HGNC:492; ANK1.
DR   MIM; 182900; -.
DR   GO; GO:0016323; C:basolateral plasma membrane; NAS.
DR   GO; GO:0008093; F:cytoskeletal adaptor activity; TAS.
DR   GO; GO:0019899; F:enzyme binding; TAS.
DR   GO; GO:0030507; F:spectrin binding; NAS.
DR   GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
DR   GO; GO:0045199; P:maintenance of epithelial cell polarity; TAS.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR011029; DEATH_like.
DR   InterPro; IPR000906; DEATH.
DR   Pfam; PF00023; Ank; 23.
DR   Pfam; PF00531; Death; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 20.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
KW   3D-structure; Alternative splicing; ANK repeat; Cytoskeleton;
KW   Direct protein sequencing; Disease mutation; Elliptocytosis;
KW   Lipoprotein; Phosphorylation; Polymorphism; Repeat.
FT   INIT_MET      0      0
FT   DOMAIN        1    826       89 kDa domain; anion exchange protein
FT                                binding domain.
FT   DOMAIN      827   1381       62 kDa domain; spectrin binding domain.
FT   DOMAIN     1382   1880       55 kDa regulatory domain; regulates the
FT                                binding of ankyrin to spectrin and the
FT                                band 3 protein.
FT   REPEAT       43     72       ANK 1.
FT   REPEAT       76    105       ANK 2.
FT   REPEAT      109    138       ANK 3.
FT   REPEAT      142    171       ANK 4.
FT   REPEAT      173    200       ANK 5.
FT   REPEAT      204    233       ANK 6.
FT   REPEAT      237    266       ANK 7.
FT   REPEAT      270    299       ANK 8.
FT   REPEAT      303    332       ANK 9.
FT   REPEAT      336    365       ANK 10.
FT   REPEAT      369    398       ANK 11.
FT   REPEAT      402    431       ANK 12.
FT   REPEAT      435    464       ANK 13.
FT   REPEAT      468    497       ANK 14.
FT   REPEAT      501    530       ANK 15.
FT   REPEAT      534    563       ANK 16.
FT   REPEAT      567    596       ANK 17.
FT   REPEAT      600    629       ANK 18.
FT   REPEAT      633    662       ANK 19.
FT   REPEAT      666    695       ANK 20.
FT   REPEAT      699    728       ANK 21.
FT   REPEAT      732    761       ANK 22.
FT   REPEAT      765    794       ANK 23.
FT   DOMAIN     1402   1486       Death.
FT   VARSPLIC   1512   1873       Missing (in isoform 2).
FT                                /FTId=VSP_000264.
FT   VARSPLIC   1874   1874       H -> D (in isoform 2).
FT                                /FTId=VSP_000265.
FT   VARSPLIC   1849   1880       TVEGPLEDPSELEVDIDYFMKHSKDHTSTPNP -> ELRGS
FT                                GLQPDLIEGRKGAQIVKRASLKRGKQ (in isoform
FT                                3).
FT                                /FTId=VSP_000266.
FT   VARIANT      20     20       R -> T.
FT                                /FTId=VAR_000595.
FT   VARIANT     462    462       V -> I (in HS).
FT                                /FTId=VAR_000596.
FT   VARIANT     618    618       R -> H (in Brueggen).
FT                                /FTId=VAR_000597.
FT   VARIANT     749    749       V -> A.
FT                                /FTId=VAR_000598.
FT   VARIANT     844    844       D -> E.
FT                                /FTId=VAR_000599.
FT   VARIANT    1285   1285       E -> D.
FT                                /FTId=VAR_000600.
FT   VARIANT    1391   1391       S -> T.
FT                                /FTId=VAR_000601.
FT   VARIANT    1591   1591       D -> N (in Duesseldorf).
FT                                /FTId=VAR_000602.
FT   VARIANT    1698   1698       R -> D.
FT                                /FTId=VAR_000603.
FT   CONFLICT    229    229       A -> S (in Ref. 2).
FT   CONFLICT   1545   1545       V -> I (in Ref. 2).
SQ   SEQUENCE   1880 AA;  206145 MW;  1C5F5E7EFD1CD428 CRC64;

Query Match             9.1%;
Best Local Similarity  26.0%;    Score 390.5;  DB 1;  Length 1880;
Matches  174;  Conservative 100;  Mismatches 246;  Indels 149;  Gaps  31;

Qy   194 RIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVK------GA----  241
Db   255 RLLLDRGAQIETKTKDEL--------------------TPLHCAARNGHVRISEILLDHGAPIQA  299

Qy   242 --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT  298
Db   300 KTKNGLSPIHMAAQGDHLD--CVRLLQYDABIDDI----TLD--HLTPLHVAAHCG  348

Qy   299 GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK  357
Db   349 HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL  404

Qy   358 NPLHHA--VEHLPPV--ILKGVMDHVKNSSEFQDLVNDPDYF---GNT-IAHYAVKNK  407
Db   405 TPLHVASFMGHLPIVKNLLQRGASPNVSNVK------VETPLHMAARAGHTEVAKYLLQNK  459

Qy   408 NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP  467
Db   460 ------------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP  505

Qy   468 LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK  525
Db   506 LHIAAREGHVETVLALL-EKEASQACMTKGFTPLHVAAKYGKVRVAELLLERDAHPNAA  564

Qy   526 GEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE  585
Db   565 GKN-----------GLTPLHVAVHHNNLDIVKL------  586

Qy   586 LVNLGDFESENILGSPNAKFPLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP  645
Db   587 LLPRG-------GSP------HSPAWN-GYTPLHIAAKQNQEVARSLLQY-GGSANAE  630

Qy   646 SERGYNAMQVAALFGDKRAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG  705
Db   631 SVQGVTPLHLAAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHVAQEGHVPVADVLI-  687

Qy   706 QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG  763
Db   688 KHGVMVDATTRMG---YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ  743

Qy   764 GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL---LDRGADINLD  820
Db   744 GHTDIVTLLLKNGASPNEVSSDGTTPLAIA-----KRLGYISVTDVLKVVTDETSFVLVS  798

Qy   821 GDHNILFDQ 829
Db   799 DKHRMSFPE 807
```

RESULT 7
```
Q9GRV5
ID   Q9GRV5        PRELIMINARY;      PRT;   1913 AA.
AC   Q9GRV5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
```

```
Db   478  VGRTPLHEVAEVGDQNMLKI--MFKLRADANIHDKEDKTPVHVAAERGDTSMVESLID--  533
                :|  ||||:||  ||  ||:||| :||:|:|::: :  ||::
Qy   380  KNSSEFQDLVNDPDYFGNTIAHYA--------------------VKNKNADLTL-------  413
                   :|:|  |  |:|:
Db   534  ----KFGGSIRARTRDGSTLHIAACSGHTSTALAFLKRGVPLFMPNKKGALGLHSAAAA  589
                        :| ||| |||  ::  ::|  :||| |||: ||| :|||:||
Qy   414  -FN---MLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVN----------  458
            ||| |||  ||||: |||: ::|: |:   |||  || :|:||:
Db   590  GFNDVVKMLIARGTNVDVRTRDNYTALHVAVQSGKASVVETLLGSGADIHVKGGELGQTA  649
            ||||||:|| ||||:|||| |||  |||||| ||||:||||| |||||||||  | |
Qy   459  ----------------------SQD----------VNGDTPLHIAVEGGSMETVLAVLNQRGADV  491
                                   :|:           ||||||:||| || ::| |||:|||:|
Db   650  LHIAASLNGAESRDCAMMLLKSGGQPDVAQMDGETCLHIAARSGNKD-IMRLLLDENADS  708
           ||||| :| || || | |||| || ||||  ||| |||||:||:|: :| ||||:|||:
Qy   492  SVQNNDGVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGG  551
             :| :|  || |||:||||||||||  ||: || :||:||||||:|| |||||||:|||
Db   709  KISSKIGETPLQVAAKSCNFEAASMILKHLSEVLTQEQLKEHVNHRTNDGFTALHYAABI  768
           |:|| ||||||:|||| ||| ||:|||||||:||| ||||:||:||||||||||||||||
Qy   552  GSR-------DTFRVVRK--NYEKCHDLATIR----AALMQDRSGGELVNLGDFESENILG  599
           | |       |  | |:|  |::| ||  ||     || ||: | |||||||| ||:||||
Db   769  EQRQLHFPGEDAKLVNLLIDYGGMVEMPSLNANETAMHMAARSGNQAVLLA---MVNKIG  825
           || ||| || ||||:|:|||  |:  |  ||||  || |||| ||| |||     |||:||
Qy   600  SPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALF  659
           ||:|||  | | |||  |  || |:|||| ||||:||||||| ||| || |||:|| ||
Db   826  AGAVQIVONKQSKN-GWSPLLEA-CARGHSGVANILLKHHARIDVFDEMGRTALHLAAFN  883
            ||| || || || ||||||:  ||| | ||| ||||| || |:| ||| |||||||:|
Qy   660  GDKEAVKMLAKSAKPSDLNFKTSATPPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDK  719
           | |||| || |||:|| | |: | ||| | ||||||| | | |||| |:||:||| |||
Db   884  GHLSLVHLLLQ--HKAFVNSKSKTGAPLHLAAQHGHVKVKVNVLVQDHGAAL-EAITLDN  940
           | ||:| |||   |  ||:| | |  |||||| | ||:||:|| ||| |:| |||:|||
Qy   720  NTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEBGGDPKIVSSLLKA---  775
            ||||||:|||| |||:||  || ||||| |||| |||||||| | |||||||||||
Db   941  QTALHFAAKFGQLAVSQTLLA-LGANPNARDDKGQTPLHLAAENDFPDVVKLFLKMRNNN  999
           ||||| |||| ||| || ||   |||| |||| |||||||||  || ||||||||| ||
Qy   776  GAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKLL  811
           || |:||||||| |||||:||  | ||| ||||||
Db   1000 RSVLTAIDHNGFTCAHIAAM----KGSLAVRELM  1030
           ||||:||||||| |||||||    |||||||||
```

RESULT 8

Q61302
ID   Q61302   PRELIMINARY;   PRT;   1848 AA.
AC   Q61302;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Erythroid ankyrin.
GN   Name=Ank1; Synonyms=Ank-1;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Neural;
RX   MEDLINE=93252825; PubMed=8486643;
RA   Birkenmeier C.S., White R.A., Peters L.L., Hall E.J., Lux S.E.,
RA   Barker J.E.,
RT   "Complex patterns of sequence variation and multiple 5' and 3' ends
RT   are found among transcripts of the erythroid ankyrin gene.";
RL   J. Biol. Chem. 268:9533-9540(1993).
DR   EMBL; X69063; CAA48801.1; -.
DR   PIR; S37771; S37771.
DR   HSSP; P16157; 1N11.
DR   MGD; MGI:88024; Ank1.
DR   GO; GO:0016323; C:basolateral plasma membrane; ISS.
DR   GO; GO:0008093; F:cytoskeletal adaptor activity; ISS.
DR   GO; GO:0019899; F:enzyme binding; ISS.
DR   GO; GO:0030507; F:spectrin binding; ISS.
DR   GO; GO:0005200; F:structural constituent of cytoskeleton; ISS.
DR   GO; GO:0045199; P:maintenance of epithelial cell polarity; ISS.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000488; Death.

DE   Hypothetical protein Y71A12B.4.
GN   Name=trp-4; ORFNames=Y71A12B.4;
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Sulston J.E.;
RL   Submitted (AUG-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL132902; CAC14420.1; -.
DR   HSSP; P16157; 1N11.
DR   WormBase; WBGene00006616; trp-4.
DR   WormPep; Y71A12B.4; CE26238.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005262; F:calcium channel activity; IEA.
DR   GO; GO:0006816; P:calcium ion transport; IEA.
DR   GO; GO:0006812; P:cation transport; IEA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR002153; Trans_receptor.
DR   Pfam; PF00023; Ank; 28.
DR   PRINTS; PR01415; ANKYRIN.
DR   PRINTS; PR01097; TRNSRECEPTRP.
DR   SMART; SM00248; ANK; 28.
DR   PROSITE; PS50088; ANK_REPEAT; 20.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   ANK repeat; Hypothetical protein.
SQ   SEQUENCE   1913 AA; 211429 MW; D421B7F253C42A27 CRC64;

```
Query Match             9.0%;     Score 387;   DB 2;   Length 1913;
Best Local Similarity  21.8%;     Pred. No. 2.7e-12;
Matches 217;   Conservative 160;   Mismatches 383;   Indels 236;   Gaps 42;

Qy     3  GDIQDQSQQDQEQDQQGAVGGAVG------NSPIERERVAAPESED---LYTVIIPK   52
          ||||||| || | |||| ||| || |    |   |:| | ||||     |  |:|||
Db    84  GENRPKSADEIKRRRRSG--GGILSRGLREMNKMVEELEQASEEPSTRKGILGTALKDM  141

Qy    53  GKRTAAPILERKSPTPE-PKVEDDEDLPPTL---------PPRTFSGEGYDDVGVSMPT  101
          | | | |  |  | |:| |  |: : |  ||         || | || |  ||    |
Db   142  EGTTYQKIYRKREETPKRSRSFDDQEMSNRVGMIEHLLRDKDPLELQQLGTDL---LTT  198

Qy   102  VSRGIYQPPIVQDSNLYSSIGG------VPQEAQYDAAA---RAGGPRKFLYGPYTFSNG  152
          | |  :|| || | | || ||      || | ||:      |||| |||  || |:|
Db   199  DTIPTDRPPL-RRSSTHLQIGKNSRIIFVPKQPSRDSVTPPDRLLGKPLFRESLTSHASS  257

Qy   153  QEIMDFEFDTPWPDVRNAVLGNKEIKEEWLTTSGPVRDI-------ADR----IV  196
          || | |  |  |  |  |:|||  || ||  ||  | |        || |
Db   258  HEEMSSE-DLAMADPQTKILYFAK-RDEWANVESEIETIKRSDFSMADNHGFTAFLLAVK  315

Qy   197  ASKGDLSEDQVE-------------------EILDIIFMNESEIAEGISNP---  228
          | |||  | |:                    || | | || ||  ||
Db   316  AGKDQIVDKMIRKGARVDYSTKDGRNATHIAAMYSGVETLELILKRYSELLRKGAGPKKQ  375

Qy   229  --LHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDL--------GAN  277
            | || ||  |  || | |  |:| |||:||  |:|| |||:||||         ||
Db   376  LAIHVACERKSKKAFPIVKRILE----DTDQRM------AEDGDGSLPIHLAFKFGNVN  424

Qy   278  AYNVL-----------DSEGNLPLHHAAKNCTGDKLKLCMEKTKTDF---IDTANFAN-  321
          |::|            :|| |||| ||| | ||  |||  |  |||    |||  ||
Db   425  IVELLLSGPSDEQTRKADGNGDTLHLAAR----SGN----TEAVRTAIAAGCDNANVQNR  477

Qy   322  --QSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHV  379
            :||| | : | |  : :|::|| | | || | ||: |  |:| ||| | : :: ||
```

```
DR   InterPro; IPR011029; DEATH_like.
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00023; Ank; 23.
DR   Pfam; PF00531; Death; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 23.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 20.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
KW   ANK repeat.
SQ   SEQUENCE   1848 AA;  202576 MW;  5744BECBF9EBA056 CRC64;

Query Match             8.9%;
Best Local Similarity   26.2%;
Matches  180;  Conservative  98;   Score 382.5;  DB 2;  Length 1848;
                              Mismatches 259;   Pred. No. 4.5e-12;  Indels 151;  Gaps  33;

Qy  221 IAEGISNPLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDIGANAYN 280
            : :|| : |: : |||::| ::|| :   :|        :|:
Db  156 LQQGHENVV-AHLINYGTKG-KVRLPALHIAARND-DTRTAAVL--LQNDPNP-----D 204

Qy  281 VLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQ---SPLHIITQKPDCSVL 337
        |:       ||: |             ||  |:           :|||:
Db  205 VLSKTGFTPLHIAAHYENLNVAQLLNRGA-----SVNFTPQNGITPLHIASRRGNVIMV 259

Qy  338 DI------------EEFT-----SRN---------LDFG-----LVDGDGKNPLHHA-- 363
        |:            ||      ||          ||        :|||| || ||
Db  260 RLLLDRGAQIETRTKDELTPLHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHMAAQG 319

Qy  364 ----------------VEHLPPVILKGVMDHVKNSSEFQDLVNDPD---YFGNTIA 400
                        ::|| :| || ||  :        |   :   :|
Db  320 DHLDCVRLLLQYNAEIDDITLDHLTPLHVAAHCGHHRVAKVLLDKGAKPNSRALNGFTPL 379

Qy  401 HYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQ 460
        | | :|:     : :|| :|| ::  ::: | :|:|||:||| ||  ||:::||||:
Db  380 HIACKKNH--IRVMELLLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVS 437

Qy  461 DVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--G 518
        :| | |||:|| |:|: | | | |:: ||||  |:| | ||:::|||  :  : :|   |
Db  438 NVKVETPLHMAARAGHTE-VAKYLLQNKAKANAKADDQTPLHCAARIGHTGMVKLLLLEN 496

Qy  519 SAKPNI-------------KGEDTVAKSLLMED--------YKGFTPLHFVAGGGSRDTF 557
          ::|               ||  |:||:||          |||||||:|| | | |
Db  497 GASPNLATTAGHTPLHTAAREGHVDTALALLEKEASQACMTKKGFTPLHVAAKYGK---V 553

Qy  558 RVVRKNYEKCHD-----------LATIRAALMQDRSGGGELVNLGDFESENIL---GSPNAK 604
        |:                     | |  :   :|                    |||
Db  554 RLAELLLE--HDAHPNAAGKNGLTPLHVAVHHNNL--DIVKL------LLPRGGSP--- 599

Qy  605 FLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEA 664
         : :   ::| :||              :|| :||: :||:   |  :  |: | |::
Db  600 ---HSPAWN-GYTPLHIAAKQNQIEVARSLLQY-GGSANAESVGGVTPLHLAAQEGHTEM 654

Qy  665 VKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHG--IDINQRMGSDKNTV 722
        |::|    : ::| | :  |||| : : : |::|   :|:  :|   ||:
Db  655 VALLL--SKQANGNLGNKSGLTPLHLHLVSQEGHVPVADVLI-KHGVTVDATTRMG---YTP 708

Qy  723 LHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEBGGDPKIVSSLLKAGAVVNRL 782
        || |   || : |::|| | ||||  |:: | |||| ||  | | ||| |||||  ||
Db  709 LHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQGHTDIVTLLLKNGASPNEV 767

Qy  783 DDNGRSVLSSAIVPGRKEKGVLGIVNKL 810
         :||  ::   ::| |:|  |: |:  |
Db  768 SSNGTTPLAIA-----KRLGYISVTDVL 790
```

RESULT 9
ANK1_MOUSE

```
ID   ANK1_MOUSE   STANDARD;   PRT;   1862 AA.
AC   Q02357;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
GN   Name=Ank1; Synonyms=Ank-1;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Erythrocyte;
RX   MEDLINE=92345717; PubMed=1386265;
RA   White R.A., Birkenmeier C.S., Peters L.L., Barker J.E., Lux S.E.;
RT   "Murine erythrocyte ankyrin cDNA: highly conserved regions of the
RT   regulatory domain.";
RL   Mamm. Genome 3:281-285(1992).
CC   -!- FUNCTION: Attach integral membrane proteins to cytoskeletal
CC       elements; bind to the erythrocyte membrane protein band 4.2, to
CC       Na-K ATPase, to the lymphocyte membrane protein GP85, and to the
CC       cytoskeletal proteins fodrin, tubulin, vimentin and desmin.
CC       Erythrocyte ankyrins also link spectrin (beta chain) to the
CC       cytoplasmic domain of the erythrocytes anion exchange protein;
CC       they retain most or all of these binding functions.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic surface of erythrocytic plasma
CC       membrane.
CC   -!- PTM: Regulated by phosphorylation (By similarity).
CC   -!- PTM: Acylated by palmitic acid group(s) (By similarity).
CC   -!- SIMILARITY: Contains 23 ANK repeats.
CC   -!- SIMILARITY: Contains 1 death domain.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M84756; AAA37236.1; -.
DR   PIR; I49502; I49502.
DR   HSSP; P16157; 1N11.
DR   MGD; MGI:88024; Ank1.
DR   GO; GO:0016323; C:basolateral plasma membrane; ISS.
DR   GO; GO:0008093; F:cytoskeletal adaptor activity; ISS.
DR   GO; GO:0019899; F:enzyme binding; ISS.
DR   GO; GO:0030507; F:spectrin binding; ISS.
DR   GO; GO:0005200; F:structural constituent of cytoskeleton; ISS.
DR   GO; GO:0045199; P:maintenance of epithelial cell polarity; ISS.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000488; Death.
DR   InterPro; IPR011029; DEATH_like.
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00023; Ank; 23.
DR   Pfam; PF00531; Death; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 23.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 20.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1
KW   ANK repeat; Cytoskeleton; Lipoprotein; Phosphorylation; Repeat.
FT   DOMAIN        1    827       89 kDa domain; anion exchange protein
FT                                binding domain.
FT   DOMAIN      828   1386       62 kDa domain; spectrin binding domain.
FT   DOMAIN     1387   1862       55 kDa regulatory domain; regulates the
FT                                binding of ankyrin to spectrin and the
FT                                band 3 protein.
FT   REPEAT       40     69       ANK 1.
FT   REPEAT       73    102       ANK 2.
FT   REPEAT      106    135       ANK 3.
FT   REPEAT      139    168       ANK 4.
```

```
FT   REPEAT      170    197       ANK 5.
FT   REPEAT      201    230       ANK 6.
FT   REPEAT      234    263       ANK 7.
FT   REPEAT      267    296       ANK 8.
FT   REPEAT      300    329       ANK 9.
FT   REPEAT      333    362       ANK 10.
FT   REPEAT      366    395       ANK 11.
FT   REPEAT      399    428       ANK 12.
FT   REPEAT      432    461       ANK 13.
FT   REPEAT      465    494       ANK 14.
FT   REPEAT      498    527       ANK 15.
FT   REPEAT      531    560       ANK 16.
FT   REPEAT      564    593       ANK 17.
FT   REPEAT      597    626       ANK 18.
FT   REPEAT      630    659       ANK 19.
FT   REPEAT      663    692       ANK 20.
FT   REPEAT      696    725       ANK 21.
FT   REPEAT      729    758       ANK 22.
FT   REPEAT      762    791       ANK 23.
FT   DOMAIN     1399   1483       Death.
SEQUENCE  1862 AA;  204242 MW;  AE6B85B5B29001E5 CRC64;
```

```
Query Match            8.9%;
Best Local Similarity  26.0%;
Matches     179;  Conservative  101;  Mismatches  257;  Indels  151;  Gaps  33;

Score 381.5;  DB 1;  Length 1862;
Pred. No. 5.1e-12;
```

```
Qy    221  IAEGISNPLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYN  280
                :    ::|| : ||:  |||           :: : || :: |       ::||
Db    148  LQQGHENVV-AHLINYGTKG-KVRLPALHIAARND-DTRTAAVL-LQNDPNP------D  196

Qy    281  VLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQ---SPLHIITQKPDCSVL  337
                ||    |||||||               :  ||:       : ||      :: :
Db    197  VLSKTGFTPLHIAAHYENLNVAQLLLNRGA------SVNFTPQNGITPLHIASRRGNVIMV  251

Qy    338  DI----------EEFT------SRN------LDFG----LVDGDGKNPLHHA----     363
                 :                    : :||       || ||  | :||:|:   :
Db    252  RLLLDRGAQIETRTKDELTPLHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHMAAQG  311

Qy    364  ---------------VEHLPPVILKGVMDHVKNSSEFQDLVNPD---YFGNTIA     400
                                ::|| ||||   : || :  | :   : : : :
Db    312  DHLDCVRLLQYNAEIDDITLDHLTPLHVAAHCGHRVAKVLLDKGAKPNSRALNGFTPL    371

Qy    401  HYAVKNKNADILTLFNMLKASGADLNVRNVGRAPIHVASSNGKANAVSGLVSCGIDVNSQ  460
                 | :     |:   :: | : |:|  ||         |:||||
Db    372  HIACKKNH--IRVMELLLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVS  429

Qy    461  DVNGDTPLHIIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--G  518
                ||  |||||:  ||| :    ::    || :  | :|| :|     ::|| | ::: |
Db    430  NVKVETPLHMAARAGHTE-VAKYLLQNKAKANAKADDQTPIHCAARIGHTGMVKLLLLEN  488

Qy    519  SAKPNI-------------KGEDTVAKSLLMED------YKGFTPLHFVAGGGSRDTF     557
                :||:|              ::|            |      |||||| ||||||
Db    489  GASPNLATTAGHTPLHIAAREGHVDTALALLEKEASQACMTKKGFTPLHVAAKYGK---V  545

Qy    558  RVVRKNYEKCHD-----------LATIRAALMQDRSGGELVNLGDFESENII---GSPNAK  604
                : :                    || |||   :  |  : |           |||
Db    546  RLAELLLE--HDAHPNAAGKNGLTPLHVAVHHNNL--DIVKL------ILPRGGSP---  591

Qy    605  FLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEA  664
                | : :||  :    || ||:|| ||: :: ||  |||:| :| :   : | ||||
Db    592  ---HSPAWN-GVTPLHIAAKQNQIEVARSLLQY-GGSANAESVQGVTPLHLAAQEGHTEM  646

Qy    665  VKMLAKSAKPSDLNFKTSATPTPLNLACLRGRGNEVVRGLVSQHG---IDINQRMGSDKNTV  722
                | | :  | ::|  |: |:|        :|| || ::  ::   |  |       :   :
Db    647  VALLL--SKQANGNLGNKSGLTPLHLAVSQEG-HVLVADVLIKHGVTVDATTRMG---YTP  700

Qy    723  LHYAISKGDSFLVQKILAHTGVDVVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRL  782
                || | :|||   |:  |:|  | : :|| | | ||||| |  |:| | |::||| :|::
Db    701  LHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQGHTDIVTLLLLKNGASPNEV  759

Qy    783  DDNGRSVLSSAIVPGRKEKGVLGIVNKL  810
                :  :||  : | :|   ||  :  | |
Db    760  SSNGTTPLAIA-----KRLGYISVTDVL  782
```

```
RESULT 10
Q61307

ID   Q61307        PRELIMINARY;       PRT;  1943 AA.
AC   Q61307; O08866; O08867; Q61305; Q61306; Q61308; Q61309; Q61310;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Ankyrin 3 (Ankyrin G) (Epithelial ankyrin) (Ankyrin-3).
GN   Name=Ank3;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1-6).
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=95340633; PubMed=7615634; DOI=10.1083/jcb.130.2.313;
RA   Peters L.L., John K.M., Lu F.M., Eicher E.M., Higgins A., Yialamas M.,
RA   Turtzo L.C., Otsuka A.J., Lux S.E.;
RT   "Ank3 (epithelial ankyrin), a widely distributed new member of the
RT   ankyrin gene family and the major ankyrin in kidney, is expressed in
RT   alternatively spliced forms, including forms that lack the repeat
RT   domain.";
RL   J. Cell Biol. 130:313-330(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 7 AND 8).
RC   STRAIN=C3H/HEJ; TISSUE=Bone Marrow;
RX   MEDLINE=97213781; PubMed=9060470; DOI=10.1083/jcb.136.5.1059;
RA   Hoock T.C., Peters L.L., Lux S.E.;
RT   "Isoforms of ankyrin-3 that lack the NH2-terminal repeats associate
RT   with mouse macrophage lysosomes.";
RL   J. Cell Biol. 136:1059-1070(1997).
CC   -!- FUNCTION: ANKYRINS ATTACH INTEGRAL MEMBRANE PROTEINS TO CYTO-
CC       SKELETAL ELEMENTS; THEY BIND TO THE ERYTHROCYTE MEMBRANE PROTEIN
CC       BAND 4.2, TO NA-K ATPASE, TO THE LYMPHOCYTE MEMBRANE PROTEIN GP85,
CC       AND TO THE CYTOSKELETAL PROTEINS FODRIN, TUBULIN, VIMENTIN AND
CC       DESMIN. ERYTHROCYTE ANKYRINS ALSO LINK SPECTRIN (BETA CHAIN) TO
CC       THE CYTOPLASMIC DOMAIN OF THE ERYTHROCYTES ANION EXCHANGE PROTEIN;
CC       THEY RETAIN MOST OR ALL OF THESE BINDING FUNCTIONS.
CC   -!- SUBCELLULAR LOCATION: PLASMA-MEMBRANE ASSOCIATED. ISOFORMS 7 AND 8
CC       ARE ASSOCIATED WITH THE LYSOSOMAL MEMBRANE.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=8;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=Q61307-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q61307-2; Sequence=VSP_050350, VSP_050353;
CC       Name=3;
CC         IsoId=Q61307-3; Sequence=VSP_050350, VSP_050353, VSP_050355;
CC       Name=4;
CC         IsoId=Q61307-4; Sequence=VSP_050354, VSP_050355;
CC       Name=5;
CC         IsoId=Q61307-5; Sequence=VSP_050352;
CC       Name=6;
CC         IsoId=Q61307-6; Sequence=VSP_050354;
CC       Name=7;
CC         IsoId=Q61307-7; Sequence=VSP_050349, VSP_050351, VSP_050355;
CC       Name=8;
CC         IsoId=Q61307-8; Sequence=VSP_050349, VSP_050351;
CC   -!- TISSUE SPECIFICITY: MAJOR FORM OF ANKYRIN IN KIDNEY AND OTHER
CC       EPITHELIAL TISSUES. ALSO EXPRESSED IN MACROPHAGES, MEGAKARYOCYTES,
CC       LEYDIG CELLS, CARDIAC, SMOOTH AND SKELETAL MUSCLE, INITIAL
CC       SEGMENTS OF AXONS, AND NODES OF RANVIER.
CC   -!- SIMILARITY: CONTAINS 24 ANK REPEATS.
DR   EMBL; L40631; AAB01603.1; -.
DR   EMBL; L40632; AAB01605.1; -.
DR   EMBL; U89275; AAB58381.1; -.
DR   EMBL; L40631; AAB01604.1; -.
DR   EMBL; L40632; AAB01607.1; -.
DR   EMBL; U89274; AAB58380.1; -.
```

```
DR   EMBL; L40632; AAB01606.1; -.
DR   EMBL; L40631; AAB01602.1; -.
DR   PIR; T42713; T42713.
DR   PIR; T42714; T42714.
DR   PIR; T42715; T42715.
DR   PIR; T42716; T42716.
DR   HSSP; P16157; 1N11.
DR   GO; GO:0045202; C:synapse; IDA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000488; Death.
DR   InterPro; IPR011029; DEATH_like.
DR   InterPro; IPR000906; DEATH_like.
DR   Pfam; PF00023; Ank; 24.
DR   Pfam; PF00531; Death; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 22.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 21.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
KW   ANK repeat; Alternative splicing; Cytoskeleton; Multigene family;
KW   Repeat.
```

```
Query Match             8.8%;  Score 378.5;  DB 2;  Length 1943;
Best Local Similarity  25.7%;  Pred. No. 8e-12;
Matches 175;  Conservative 98;  Mismatches 275;  Indels 133;  Gaps 27;

Qy  184 TSGPVRDIADRIVASKGDLSEDQVEEILDIFMNES----EIAEGISNPLH------ 230
            ||||||  |||   : |:: |||
Db  180 TKGKVRLPALHIAARKDD------TKAAALLQNDTNADVESKSGFT-PLHIAAHYGNIN 232

Qy  231 ------------ADVDNNPVKGAKNVMTLMHIVYACDVDPRIVKALGEVENDEGDLGANAYNV 281
                         |:|: :|         :|: ||
Db  233 VATLLLNRAAAVDFT---ARNDITPLH-VASKRGNANMVKLL------LDRGAK-IDA 279

Qy  282 LDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIBE 341
            |||| ||| : :                          |||  :
Db  280 KTRDGLTPLHCGARSGHEQVVEMLLDRSAPILSKTKN--GLSPLHMATQGDHLNCVQL- 335

Qy  342 FTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVQNSSEFQDLVNDPDYFGNTIAH 401
              : |||   :   |||   |                          ||||
Db  336 LLQHNVPVDDVTNDYLTALHVAAHCGHYKVAKVLLD--KKASPNAKALN------GFTPLH 388

Qy  402 YAVKNKNADLTLFMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQD 461
            ||  |         |     ||   ||| :
Db  389 IACK-KNR-IRVMELLLKHGASIQAVTESGLTPIHVAAFMGHVNIVSQLMHHGASPNTTN 446

Qy  462 VNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GS 519
            | | ||| :|        |    :| |      | | ||  :||| |||     |:
Db  447 VRGETALHMAARSGQAE-VVRYLVQDGAQVEAKAKDDQTPLHISARLGKRADIVQQLLQQG 505

Qy  520 AKPNI----------------------------KSLLMEDYKGFTPLHFVAGGSRDTFR 558
            | ||                              :| |||||||
Db  506 ASPNAATTSGYTPLHLAAREGHEDVAAFLLDHGASLSITTKKGFTPLHVAAKYGKLEV-- 563

Qy  559 VVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSP 618
            |:                                  :: :
Db  564 ------------------------ASLLLQKS----------------AGKSGLTP 584

Qy  619 ARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAKSAKPSDLN 678
            :| :                |  | :        |: |:| : :: :
Db  585 LHVAAHYDNQXVALLLLD--QGASPHAAAKNGYTPLHIAAKKNQMDIATSLLEYG--ADAN 641

Qy  679 FKTSATPTPNLACLRGDNEVVRGLVGQH-GIDINQRMGSDKNTVLHYAISKGDSFLVQK 737
             | ::  | :            |:         |:         ||||:  |||
Db  642 AVTRQGIASVHLAAQEGHVDMVSLLLSRNANVNLSNKSGL--TPLHLA-AQEDRVNVAE 697

Qy  738 ILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPG 797
            ||  |  |      ||||||||:  | | :|  || |   |  ||| || |:| :|
Db  698 VLVNQGAHVDAQTKMGYTPLHVGCHYGHGNIKIVNFLLQHSAKVNAKTKNGYTALHQA---- 753

Qy  798 RKEKGVLGIVNKLLDRGADIN 818
              :::  | |:||  |||
Db  754 -AQQGHTHIINVLLQNNASPN 773
```

```
RESULT 11
ANK3_HUMAN
ID   ANK3_HUMAN    STANDARD;    PRT;   4377 AA.
AC   Q12955;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Ankyrin 3 (ANK-3) (Ankyrin G).
GN   Name=ANK3;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain stem;
RX   MEDLINE=95138209; PubMed=7836469; DOI=10.1074/jbc.270.5.2252;
RA   Kordeli E., Lambert S., Bennett V.;
RT   "AnkyrinG. A new ankyrin gene with neural-specific isoforms localized
```

```
FT   DOMAIN        1     826    ANION EXCHANGE PROTEIN BINDING DOMAIN.
FT   DOMAIN       23     810    24 X ANK MOTIF REPEATS.
FT   DOMAIN      856    1437    SPECTRIN BINDING DOMAIN.
FT   DOMAIN     1438    1943    REGULATORY DOMAIN (REGULATES BINDING OF
FT                              ANKYRIN TO SPECTRIN AND THE BAND 3
FT                              PROTEIN).
FT   REPEAT       23      55    ANK MOTIF 1.
FT   REPEAT       56      88    ANK MOTIF 2.
FT   REPEAT       89     121    ANK MOTIF 3.
FT   REPEAT      122     154    ANK MOTIF 4.
FT   REPEAT      155     183    ANK MOTIF 5.
FT   REPEAT      184     216    ANK MOTIF 6.
FT   REPEAT      217     249    ANK MOTIF 7.
FT   REPEAT      250     282    ANK MOTIF 8.
FT   REPEAT      283     315    ANK MOTIF 9.
FT   REPEAT      316     348    ANK MOTIF 10.
FT   REPEAT      349     381    ANK MOTIF 11.
FT   REPEAT      382     414    ANK MOTIF 12.
FT   REPEAT      415     447    ANK MOTIF 13.
FT   REPEAT      448     480    ANK MOTIF 14.
FT   REPEAT      481     513    ANK MOTIF 15.
FT   REPEAT      514     546    ANK MOTIF 16.
FT   REPEAT      547     579    ANK MOTIF 17.
FT   REPEAT      580     612    ANK MOTIF 18.
FT   REPEAT      613     645    ANK MOTIF 19.
FT   REPEAT      646     678    ANK MOTIF 20.
FT   REPEAT      679     711    ANK MOTIF 21.
FT   REPEAT      712     744    ANK MOTIF 22.
FT   REPEAT      745     777    ANK MOTIF 23.
FT   REPEAT      778     810    ANK MOTIF 24.
FT   VARSPLIC      1     805    Missing (in isoform 7 and isoform 8).
FT                              /FTId=VSP_050349.
FT   VARSPLIC      1     849    Missing (In isoform 2 and isoform 3).
FT                              /FTId=VSP_050350.
FT   VARSPLIC    833     853    Missing (in isoform 7 and isoform 8).
FT                              /FTId=VSP_050351.
FT   VARSPLIC    834     855    VRKASAPEKLSDGEYISDGEEG -> GDKCTWFKIPKVQEV
FT                              LVK (in isoform 5).
FT                              /FTId=VSP_050352.
FT   VARSPLIC    850     855    SDGEEG -> MALPHS (in isoform 2 and
FT                              isoform 3).
FT                              /FTId=VSP_050353.
FT   VARSPLIC    852     852    G -> GDKCTWFKIPKVQEVLVKS (in isoform 4
FT                              and isoform 6).
FT                              /FTId=VSP_050354.
FT   VARSPLIC   1569    1764    Missing (in isoform 3, isoform 4 and
FT                              isoform 7).
FT                              /FTId=VSP_050355.
SQ   SEQUENCE   1943 AA;  212013 MW;  B23C4B6F82904804 CRC64;
```

RT   at the axonal initial segment and node of Ranvier.";
RL   J. Biol. Chem. 270:2352-2359(1995).
CC   -!- FUNCTION: Membrane-cytoskeleton linker. The neural-specific
CC   isoforms may participate in the maintenance/targeting of ion
CC   channels and cell adhesion molecules at the nodes of Ranvier and
CC   axonal initial segments.
CC   -!- SUBUNIT: Neural-specific isoforms may be a constituent of a
CC   neurofascin/NRCAM/ankyrin G complex.
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=1;
CC   Comment=A number of isoforms are produced;
CC   Name=1; Synonyms=480-kDa isoform;
CC   IsoId=Q12955-1; Sequence=Displayed;
CC   -!- TISSUE SPECIFICITY: Expressed in brain and other tissues. Isoform
CC   1 is neural-specific.
CC   -!- SIMILARITY: Contains 23 ANK repeats.
CC   -!- SIMILARITY: Contains 1 death domain.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; U13616; AAA64834.1; -.
DR   PIR; A55575; A55575.
DR   HSSP; P16157; 1N11.
DR   Genew; HGNC:494; ANK3.
DR   MIM; 600465; -.
DR   GO; GO:0006605; P:protein targeting; NAS.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000488; Death.
DR   InterPro; IPR011029; DEATH_like.
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00023; Ank; 24.
DR   Pfam; PF00531; Death; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   TIGRFAMs; TIGR01784; T_den_put_tspse; 1.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 21.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
KW   Alternative splicing; ANK repeat; Cytoskeleton; Repeat.
FT   REPEAT   73   102   ANK 1.
FT   REPEAT   106   135   ANK 2.
FT   REPEAT   139   168   ANK 3.
FT   REPEAT   172   201   ANK 4.
FT   REPEAT   203   230   ANK 5.
FT   REPEAT   234   263   ANK 6.
FT   REPEAT   267   296   ANK 7.
FT   REPEAT   300   329   ANK 8.
FT   REPEAT   333   362   ANK 9.
FT   REPEAT   366   395   ANK 10.
FT   REPEAT   399   428   ANK 11.
FT   REPEAT   432   461   ANK 12.
FT   REPEAT   465   494   ANK 13.
FT   REPEAT   498   527   ANK 14.
FT   REPEAT   531   560   ANK 15.
FT   REPEAT   564   593   ANK 16.
FT   REPEAT   597   626   ANK 17.
FT   REPEAT   630   659   ANK 18.
FT   REPEAT   663   692   ANK 19.
FT   REPEAT   696   725   ANK 20.
FT   REPEAT   729   758   ANK 21.
FT   REPEAT   762   791   ANK 22.
FT   REPEAT   795   825   ANK 23.
FT   DOMAIN   1519   1898   Ser-rich.
FT   DOMAIN   4090   4174   Death.
SQ   SEQUENCE   4377 AA; 480399 MW; F42379E55768B684 CRC64;

Best Local Similarity   25.5%; Pred. No. 3e-11;
Matches 175; Conservative   92; Mismatches 277; Indels 141; Gaps 26;

```
Qy   184  TSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNES-----BIAEGISNPLH--------  230
               |||||||   |      ::     | :       :::   ||||
Db   197  TKGKVRLPALHIAARKDD------TKAAALLQNDNNADVESKSGFT-PLHIAAHYGNIN  249

Qy   231  -------ADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNV  281
                       : ||||  |    :|   |   |   ||
Db   250  VATLLLNRAAAVDFT---ARNDITPLH-VASKRGNANMVKLL------LDRGAK-IDA  296

Qy   282  LDSEGNLPLHHAAKNCTGDKLKLCMBKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEB  341
               ||||  |  |  :         :   : |      ::|||
Db   297  KTRDGLTPLHCGARSGHEQVVEMLLDRAAPILSKTKN--GLSPLHMATQGGDHLNCVQL--  352

Qy   342  FTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAH  401
               | :        ::       :       ::   :|   : |     ||  ||
Db   353  LLQHNVPVDDVTNDYLTALHVAAHCGHYKVAKVLLDKKANP----------NAKALNGFTPLH  405

Qy   402  YAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQD  461
               :|    ::  |  ||| | ||  | |  |  ||    |   |  ::|:  |  ||
Db   406  IACK-KNR-IKVMELLLKHGASIQAVTESGLTPIHVAAFMGHVNIVSQLMHHGASPNTTN  463

Qy   462  VNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GS  519
               |||||||||    | ::    |:|||  | :||||| ||  || ||  :  |   |
Db   464  VRGETALHMAARSGQAE-VVRYLVQDGAQVRAKAKDDQTPLHISARLGKADIVQQLLQQG  522

Qy   520  AKPNI------------------------KGEDTVA------KSLLMEDYKGFTPLHFVAGGGSRDTFR  558
               ||:|                        || |         ::|||:| |||| |
Db   523  ASPNAATTSGYTPLHLSAREGHEDVAAFLLDHGASLSITTKKGFTPLHVAA---------  573

Qy   559  VVRKNYEKCHDLATTRAALMQDRSGGELVNLGDFESENIL----GSPNAKFLQHIQSANF  614
               :  |   :|     |     | :      ||  :|        | : |
Db   574  --------------------------------KVGKLEVANLLLQKSASPDA---------AGKS  597

Qy   615  GFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEAVRMLAKSAKP  674
               | ||:   |    : |  :|      | ||       : | |  | |::  ||
Db   598  GLTPLHVAAHYDNQKVALLLD-QGASPHAAAKNGYTPLHIAAKKNQMDIATTLLEYG--  654

Qy   675  SDLNFKTSATPTPLNLACLRGDNEVVRGLVGQH-GIDINQRMGSDKNTVLHYAISKGDSF  733
               :|  ||:| | |    |  ||| : |  :    | |:   | :  :||
Db   655  ADANAVTRQGIASVHLAAQEGHVDMVSLLLGRNANVNLSNKSGL---TPLHLA-AQEDRV  710

Qy   734  LVQKIIAHTGVDVNCENNLGQTPLHIAVEGGDPKIVSSLLKAGAVVNRLDDNGRSVLSSA  793
               ::  :      |  :|  || ||:|||:  |   | |::  |  | | :|
Db   711  NVAEVLVNQGAHVDAQYKMGYTPLHVCHYGNIKIVNFLLQHSAKVNAKTKNGYTPLHQA  770

Qy   794  IVPGRKEKGVLGIVNKLLDRGADIN  818
               | | ::| : :|::|| |
Db   771  -----AQQGHTHIINVLLQNNASPN  790
```

RESULT 12
Q9N180
ID   Q9N180   PRELIMINARY;   PRT;   1136 AA.
AC   Q9N180;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Ankyrin 1 (Fragment).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=breed Japanese Black; TISSUE=Bone marrow;
RA   Matsumoto M., Inaba M., Koshino I., Saito D., Ono K.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF222766; AAF61702.1; -.
DR   HSSP; P16157; 1N11.
DR   InterPro; IPR002110; ANK.

Query Match   8.8%; Score 377.5; DB 1; Length 4377;

```
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00023; Ank; 23.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 23.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 20.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   ANK repeat.
FT   NON_TER      1136       1136
SQ   SEQUENCE   1136 AA;  123182 MW;  07CDBB156BACB8C7 CRC64;
```

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Amygdala;
RA   Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA   Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA   Otsuki T., Sato T., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA   Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK126888; BAC86737.1; -.
DR   HSSP; Q00420; 1AWC.
DR   InterPro; IPR002110; ANK.
DR   Pfam; PF00023; Ank; 17.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 19.
DR   PROSITE; PS50088; ANK_REPEAT; 17.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   ANK repeat.
SQ   SEQUENCE    687 AA;   74096 MW;  B5169C83017FF571 CRC64;
```

```
Query Match             8.7%;  Score 373.5;  DB 2;  Length 687;
Best Local Similarity  24.0%;  Pred. No. 3.3e-12;
Matches  190;  Conservative  115;  Mismatches  299;  Indels  187;  Gaps  35;
```

```
Qy      72 VEDDEDLPPTLPPRTFSGEGYDDVGVSMPTVSRGIYQPP----IVQDSNLYSSI-GGVPQ 126
                :  ||     :        : :     :: :            :: |||::: ::|:
Db       9 LEEVEDESPAFISK-----------------------LPQENKSLHSPPSGNVLVRYPSLVQAIFNGDPD 55
```

```
Qy     127 EAQYDAAAARAGGPRKFLYGPYTPSNGQBIMDFEPDTPWPDVRNAVLGNKEIKE----- 179
                        :  ::            :| :||
Db      56 EV-------RALIFK--- KEDVNFQDNEKRTPLHAA--AYLGDAEIIELLLLSGA 98
```

```
Qy     180 --------EWLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNESEIAEGISN---PL 229
                   ||| :   :          :           :| |   |  :         :
Db      99 RVNAKDSKWLT--- PLH---- RAVASCS-------EEAVQVLLKHSADVNARDKNWQTPL 144
```

```
Qy     230 HADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLP 289
                : :|   :||   : :::       |  :                      NVSDRAGRTA 176
Db     145 HIAAANKAVKCAEALVPLLSNV--------------------------- NVSDRAGRTA 176
```

```
Qy     290 LHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDF 349
            ||| :  |||  :|    : :: ||  |:  |   :   :|| ::    ||| ::     ||
Db     177 LHHAAFSGHGEMVKLLLSRGAN--INAFKDRRAIHWAAYMGHIEVVKL--LVSHGAEV 232
```

```
Qy     350 GLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAHYAVKNKNA 409
            |: |   ||||| :|||  | |:   ::   ::  |     || :||  |: ||  ::| |
Db     233 TCKDKKSYTPLHAAAS--------SGMISVVKYLLDLGVDMNEPNAYGNTPLHVACYN-GQ 284
```

```
Qy     410 DLTLFNMLKASGADLNVRNVVGRAPIH-VASSNGKANAVSGLVSCGIDVNSQDVNGDTPL 468
                       :  ||| :   | :     :|   :    |  |:      :| :| : :|
Db     285 DVVV-NELIDCGAIVNQKNEKGFTPLHFAAASTHGALCLELLVGNGADVNMKSKDGKTPL 343
```

```
Qy     469 HIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKALGSAKPNIKGED 528
            | : : | |   : |:|   || :  |  || |: |||:|   |:  ||| |: ||
Db     344 HMTALHGRFSRSQTII-QSGAVICEDKNGNTPLHIAARYGHELLINTLITS------GAD 397
```

```
Qy     529 TVAKSLLMEDYKGFTPLHFVAGGSRDTFRVVRKNYEKCHDLATI----RAALMQDRSGG 584
                  ||   |  ||||                 | :||    |           |    :|
Db     398 TAKRGI------HGMPPLHLAALSGFSD--- CCRKLLSSGFDIDTPDDFGRTCLHAAAAGG 449
```

```
Qy     585 ELVNLGDFESENILGSPNAKFLQHIQSANPGFSPARRGIVSSNHNVMKDILN---FVGD- 640
            | ::|:|      :| ||             ::| ||  |    :  |:      ||||
Db     450 NL-------ECLNLLLNTGADF---NKKDKFGRSPLHYAAANCNYQCLFALVGSGASVNDL 500
```

```
Qy     641 ------SLH-------------------------------LPSERGYNAMQVAALFGDKEAVKMLAKS 671
                   |||
Db     501 DERGCTPLHYAATSDTDGKCLEYLLRNDANPGVRDKQGTNAVHYSAAYGHRLCLQLIA-S 559
```

```
Query Match             8.7%;  Score 377;  DB 2;  Length 1136;
Best Local Similarity  25.9%;  Pred. No. 4.4e-12;
Matches  178;  Conservative  99;  Mismatches  260;  Indels  151;  Gaps  32;
```

```
Qy     221 IAEGISNPLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYN 280
                : |   | : :              :  || : ||               :|
Db     151 LQQGHENVV-AHLINYGTKG-KVRLPALHIAARND-DTRTAAVL--LQNDPNP------D 199
```

```
Qy     281 VLSKTGFTPLHIAAHYENLNVAQLLLNRGA-----SVNFTPQNGITPLHIASRRGNVIMV 254
            |  :   ||||   |: :: :  ||| ||     :  :      |||| :|::||:|:
Db     200 VLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQ---SPLHIITQKPDCSVL 337
```

```
Qy     338 DI--------EEFT-----SRN-------LDFG-----LVDGDGKNPLHHA-- 363
               :            ||              |:            | |||||||:
Db     255 RLLLDRGAQIETRTKDELTPLHCAARNGHLRISEILLDHGAPIQAKTKNGLSPIHMAAQG 314
```

```
Qy     364 ---------VEHLPPVILKGVMDHVKNSSEFQDLVNDPD---YFGNTIA 400
                     :  :            : |   ||  ||
Db     315 DHLDCVRLLQYNAEIDDITLDHLTPLHVAAHCGHHRVAKVLLDKGAKPNSRALNGFTPL 374
```

```
Qy     401 HYAVKNKNADLITLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQ 460
            |  |    |  : |       |  : :      |:
Db     375 HIACKKNH--IRVMELLLKMGASIDAVTESGLTPLHVASFMGHPPIVKSLLQREASPNVS 432
```

```
Qy     461 DVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--G 518
            :| ||||| :|    : |   |:| ||   :  :: :| |:  |:||   |:|||:    |
Db     433 NVKVETPLHMAAARAGHTE-VAKYLLQNKAKDDQTPLHCAARIGHTNMVKLLLEN 491
```

```
Qy     519 SAKPNI-------KGEDTVAKSLLMED-------YKGFTPLHFVAGGGSRDTF 557
                   |:            |||  || |         ||||||        ::
Db     492 NANPNLATTAGHTPLHIAAREGHVETALALLEKEASQTCMTKKGFTPLHVAAKYGK---V 548
```

```
Qy     558 RVRKNYEKCHD------------LATIRAALMQDRSGGELVNLGDFESENIL---GSPNAK 604
            :::                    |  :         :           | :
Db     549 RMAELLLE--HDAHPNAAGKSGLTPLHVAVHHNL--DVVRL-------LLPRGGSPHSP 597
```

```
Qy     605 FLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEA 664
            :  |:    || : |
Db     598 ALN-------GYTPLHIAAKQNQLEVARSLLQY-GGSANAESVGGVTPLHLAAQEGHAEM 649
```

```
Qy     665 VKOMLAKSAXPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHG--IDINQRMGSDKNTV 722
               ::|   |:         |||      ::    :|  || ::   :|:  ::  :|  :
Db     650 VALL--SKQANGNLGNKSGLTPLHLVAQEGHIPVADVLI-KHGVTVDATTRMG---YTP 703
```

```
Qy     723 LHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGGDPKIVSSLLKAGAVVNRL 782
            || |:  ||        : |:     ||| :| |||   |       |:||  || :|
Db     704 LHVASHYGNIKLVKFLLQHK-ADVNAKTKLGYSPLHQAAQQGHTDIVTLLLKHGASPNEV 762
```

```
Qy     783 DDNGRSVLSSAIVPGRKEKGVLGIVNKL 810
            : ||:|:   ::|  |::|  | :|:|
Db     763 SSNGTTPLAIA------KRLGYISVTDVL 785
```

```
RESULT 13
Q6ZT57
ID   Q6ZT57          PRELIMINARY;      PRT;    687 AA.
AC   Q6ZT57;
DT   05-JUL-2004  (TrEMBLrel. 27, Created)
DT   05-JUL-2004  (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004  (TrEMBLrel. 27, Last annotation update)
DE   Hypothetical protein FLJ44940.
```

```
Query Match        8.6%;   Score 371.5;  DB 2;   Length 1726;
Best Local Similarity   25.6%;  Pred. No. 1.6e-11;
Matches 174;  Conservative  98;  Mismatches 276;  Indels 133;  Gaps  27;

Qy   184 TSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNES----------EIAEGISNPLH---------- 230
             |:|:                                        :||:
Db   180 TKGKVRLPALHIAARKDD----------TKAAALLLQNDTNADVESKSGFT-PLHIAAHYGNIN 232

Qy   231 ---------ADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNV 281
                       |||||   ||  |  |   |  ||    :||
Db   233 VATLLLNRAAAVDFT----ARNDITPLH-VASKRGNANMVKLL-------LDRGAK-IDA 279

Qy   282 LDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEE 341
             :||:|||             :::||:   :::
Db   280 KTRDGLTPLHCGARSGHEQVVEMLLDRSAPILSKTKN--GLSPLHMATQGDHLNCVQL-- 335

Qy   342 FTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAH 401
               |    | |||  |||           |||  |     :           ::| | |
Db   336 LLQHNVPVDDVTNDYLTALHVAAHCGHYKVAKVLLD--KKASPNAKALN------GFTPLH 388

Qy   402 YAVGKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQD 461
               |||||    |      |       |||||||||
Db   389 IACK-KNR-IRVMELLLKHGASIQAVTESGLTPIHVAAFPMGHVNIVSQLMHHGASPNTTN 446

Qy   462 VNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GS 519
             |||||||||        |     |  |      |||||||            | |  |
Db   447 VRGETALHMAARSGQAB--VVRYLVQDGAQVEAKADKDQTPLHISARLGKADIVQQLQQG 505

Qy   520 AKPNI---------------KGEDTVA-------KSLLMEDYKGFTPLHFVAGGSRDTFR 558
             || :                |    ||        |::  : |:|||||:|    | | |
Db   506 ASPNAATTSGYTPLHIAAREGHEDVAAFLLDHGASLSITTKKGFTPLHVAAKYGKLEV-- 563

Qy   559 VVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSP 618
                                   |:|:                       ||
Db   564 ---------------------ASLLQKS---------ASPDA--------AGKSGLTP 584

Qy   619 ARRGIVSSNHNVMKDILNFVGDSLIHLPSERGYNAMQVAALFGDKEAVKMLAKSAKPSDLN 678
             :   :|:   |:|   ||:|:     ::
Db   585 LHVAAHYDNQKVALLLLD-QGASPHAAAKNGYTPLHIAAKKNQMDIATSLLEYG--ADAN 641

Qy   679 FKTSAITPTPLNLACLRGDNEVVRGLVGQH-GIDINQRMGSDKNTVLHYAISKGDSFLVQK 737
                 |  ||||||||:||||   |  |  |   |:|:   ||  ||||      |   |  |
Db   642 AVTRQGIASVHLAAQEGHVDMVSLLLSRNANVNLSNKSGL---TPLHLA-AQEDRVNVAE 697

Qy   738 ILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPG 797
              :|:|:  |   |  |||||||        |  |  | |        | :  | ||  | |
Db   698 VLVNQGAHVDAQTKMGYTPPHVGCHYGNIKIVNFLLQHSAKVNAKTKNGYTALHQA---- 753

Qy   798 RKEKGVLGIVNKLLDRGADIN 818
               :|:   :|:| |  | | |
Db   754 -AQQGHTHINVLLQNNASPN 773
```

```
RESULT 14
Q8VC68
ID   Q8VC68             PRELIMINARY;     PRT;  1726 AA.
AC   Q8VC68;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Ank3 protein.
GN   Name=Ank3;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Salivary gland;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RL   "Generation and initial analysis of more than 15,000 full-length human
RL   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Salivary gland;
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC021657; AAH21657.1; -.
DR   HSSP; P16157; 1N11.
DR   MGD; MGI:88026; Ank3.
DR   GO; GO:0045202; C:synapse; IDA.
DR   Pfam; PF000023; Ank; 24.
DR   Pfam; PF00531; Death; 1.
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 22.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00218; ZU5; 1.
KW   ANK repeat.
DR   PROSITE; PS50088; ANK_REPEAT; 21.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
SQ   SEQUENCE   1726 AA;  188238 MW;  43014B67E00B1CA CRC64;
```

```
Qy   672 AKPSDLNFKTS---------ATPTPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDKN 720
              |:|:|||          |||:||   ||||||||   |  |  |   |:|:   ||
Db   560 ETPLDVLMETSGTDMLSDSDNRATISPLHLAAYHGHQALEVLIV-QSLLDLDVRNSSGR- 617

Qy   721 TVLHYAISKGDSFLVQKILAHTGVDVNCENN-LGQTPLHLAVEGGDPKIVSSLLKAGAVV 779
              :||                |  |    |  |   ||||||            | |
Db   618 TPLDLAAFKGHVECVD-VLINQGASIIVKDYILKRTPIHAAATNGHSECLRLLIGNAEPQ 676

Qy   780 NRL---DDNGR 787
             ||    ||:|
Db   677 NAVDIQDGNGQ 687
```

```
RESULT 15
Q9VMR4
ID   Q9VMR4             PRELIMINARY;     PRT;  1619 AA.
AC   Q9VMR4; Q9NGR8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   CG11020-PA (Mechanosensory transduction channel NOMPC).
GN   Name=nompC; ORFNames=CG11020;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
```

RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan H., Zhang Q., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfied E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RA   Walker R.G., Willingham A.T., Zuker C.S.;
RT   "A Drosophila mechanosensory transduction channel.";
RL   Science 0:0-0(2000).
DR   EMBL; AE003609; AAF52248.3; -.
DR   EMBL; AF242296; AAF59842.1; -.
DR   HSSP; P09959; 1SW6.
DR   FlyBase; FBgn0016920; nompC.
DR   GO; GO:0030506; F:ankyrin binding; NAS.
DR   GO; GO:0008381; F:mechanically-gated ion channel activity; NAS.
DR   GO; GO:0009592; P:detection of sound; IMP.
DR   GO; GO:0007638; P:mechanosensory behavior; IMP.
DR   GO; GO:0007605; P:perception of sound; IMP.
DR   GO; GO:0050974; P:sensory transduction of mechanical stimulus; TAS.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR000550; Hppk.
DR   Pfam; PF00023; Ank; 29.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 29.
DR   PROSITE; PS50088; ANK_REPEAT; 20.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
DR   PROSITE; PS00794; HPPK; UNKNOWN_1.
KW   ANK repeat.
SQ   SEQUENCE   1619 AA; 176843 MW; B4FE239AB81D1138 CRC64;

```
Query Match             8.6%;   Score 371;  DB 2;   Length 1619;
Best Local Similarity   21.9%;  Pred. No. 1.6e-11;
Matches  187;  Conservative 133;  Mismatches 299;  Indels 236;  Gaps 35;

Qy       119 SSIGGVPQEA------QYDAAARAGGPRKFLYGPYTFSNG-QEIMDFEFDTPWP------ 165
                 :: || |       :: :::|
Db        22 SSLTGPPDESATPSERATPASKADSDPK----DDSSSNGDKKDMDL-FPAPKPPSAGASI 76

Qy       166 -DVRNAVLGNKEIKEEWLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIPMNESEIAEG 224
              |    |||| || :|      : ||| | ::| |  |
Db        77 RDTANKVLG-LAMKSEWTPIEAELKKL-EKYVANVGE------------ 111

Qy       225 ISNPLHADVDNNPVKGAKNVMT-LMHLVTAC-DVDPRIVKALGEVENDEGDLGANAYNVL 282
                 || || ||   |:||        : :|| ||    |||
Db       112 ------DGNHIPLAGVHDMNTGMTPLMYATKDNKTAIMDRMIELGADVGARNNDNYNV- 163

Qy       283 DSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEE- 341
             |   | |||  |  || ||| |      |::  :|   |: |:   ||:|:|  |  ::
Db       164 ------LHIAAMYSREDVVKLLLLTKRGVDPFSTGGSRSQTAVHIVSSRQTGTATNILRA 216

Qy       342 -FTSRNLDFGL-VDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTI 399
              |  :|| | | ||  | ||||| | ||  |: |  |: |  || |  ||| || : :|
Db       217 LLAAAGKDIRLKADGRGKIPLLLAVESGNQSMCRELL----AAQTAEQLKATTANGDTA 271

Qy       400 AHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNS 459
             | |: ||| ||:::  :::|:: | || : | ||||| : ||||||: : | |    :
Db       272 LHLAARRRDVDMV--RILVDYGTNVDTQNGEGQTPLHIAAAEGDEALLKYFYGVRASASI 329

Qy       460 QDVNGDTPLHIAVEGGSMETV----------------LAVLN-----------Q 486
             |:|||||||| |  ||                     | ::|            :
Db       330 ADNQDRTPMHLAAENGHAHVIEILADKFKASIFERTKDGSTLMHIASLNGHAECATMLFK 389

Qy       487 RGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIKGEDTV----------AKSLL 535
              ||:|  | |||:||:   | ||| |::  |    :| ||||| ||          | :||
Db       390 KGVVLHMPNKDGARSIHTAAAYGHTGIINTLLQKGEKVDVTTNDYTALHIAVESAKPAV 449

Qy       536 MEDYKGFTPLHFVAGGSRDTFRVVRKNYEKCHDLATI----RAALMQDRSG------ 583
               ::  | |||| |||  ||: |||| :  :| ::|         : :|||:|
Db       450 VETLLGFGADVHVRGGKLRET------------PLHIAARVKGDRCALMLLKSGASPNLTTD 500
```

```
Qy  584 ----------GELVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVM 631
                      :|||: :|       :     : .  :  : .: .:::
Db  501 DCLTPVHVAARHGNLATLMQLLEDE--GDP-------LYKSNTGETPLHMACRACHPDIV 551

Qy  632 KDILNFVGDSLHLP---------SERGYNAM---------QVAALFGDKEAVKMLAKSA 672
        :  ::: :|---| :          |: :           ||:|:||: || ::
Db  552 RHLIETVKEK-HGPDKATTYINSVNEDGATALHYTCQITKEEVKIPESDKQIVRMLLENG 610

Qy  673 --J------------------------------KPSD----LNFKTSATPTPLNLACLR 694
                                          :|: ::|||||| |:|||
Db  611 ADVTLQTKTALETAFHYCAVAGNNDVLMEMISHMNPTDIQKAMNRQSSVGWTPLLIACHR 670

Qy  695 GDNEVVRGLVGQHG-IDINQRMGSDKNTVLHYAISKG-----DSFLVQKILAHTGVDVNC 748
        |  |: |.|| :|: :|:      | |:     .|        :|     |   |:
Db  671 GHMELVNNLLANHARVDV---FDTEGRSALHLAAERGYLHVCDALLTNKAF------INS 721

Qy  749 ENNLGQTPLHLAVEGGDPKIVSSLLK-AGAVNRLDDNGRSVLSSAIVPGRKEKGVLGIV 807
        ::.:|:|:|||| |||: ||    ||: ||:| :|  |  :     |   |   |  |
Db  722 KSRVGRTALHLAAMNGFTHLVKFLIKDHNAVIDILTLRKQTPLHLAAASGQME-----VC 776

Qy  808 NKLLDRGADINLDGD 822
        ||: ||:||:::
Db  777 QLLLELGANIDATDD 791
```

Search completed: February 26, 2005, 18:53:54
Job time : 182 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
, Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 26, 2005, 18:36:06 ; Search time 174 Seconds
(without alignments)
1847.114 Million cell updates/sec

Title: US-10-701-038-2
Perfect score: 4310
Sequence: 1 PSGDIQDSQQDQQEQDQQQ.........DRGADINLDGDHNILFDQCL 831

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4310	100.0	831	8	ADO15511	Ado15511 Immunorea
2	1963	45.5	390	5	ADG31916	Adg31916 E canis P
3	1300.5	30.2	831	8	ADO15510	Ado15510 Immunorea
4	421	9.8	1053	4	AAU28174	Aau28174 Novel hum
5	421	9.8	1059	7	ADK65850	Adk65850 Angiogene
6	421	9.8	1059	4	ADS88330	Ads88330 Human pro
7	413	9.6	705	4	AAU20496	Aau20496 Human sec
8	391.5	9.1	1880	8	ADR90358	Adr90358 Full leng
9	390.5	9.1	1881	7	ADD47763	Add47763 Human Pro
10	382.5	8.9	899	8	ABM80426	Abm80426 Tumour-as
11	382.5	8.9	1536	6	ABU11523	Abu11523 Human MDD
12	377.5	8.9	4274	6	ABG00972	Abg00972 Novel hum
13	377.5	8.8	4386	6	ABG07375	Abg07375 Novel hum
14	373.5	8.7	687	8	ADR08749	Adr08749 Human pro
15	373.5	8.7	1056	6	AAE33684	Aae33684 Human str
16	370.5	8.6	994	7	ADK65799	Adk65799 Angiogene
17	370.5	8.6	1012	7	ADK65847	Adk65847 Angiogene
18	370	8.6	1619	4	AAU00986	Aau00986 Drosophil
19	368.5	8.5	1330	7	ADM04552	Adm04552 Human pro
20	368.5	8.5	1839	7	ADD27862	Add27862 Human bra
21	368.5	8.5	1872	4	AAM79160	Aam79160 Human pro
22	368.5	8.5	2622	7	ADE55508	Ade55508 Rat Prote
23	368.5	8.5	4397	4	ABG21944	Abg21944 Novel hum
24	368	8.5	1619	4	AAU00970	Aau00970 Drosophil
25	366.5	8.5	1619	4	AAU00988	Aau00988 Drosophil
26	364.5	8.5	919	7	ADM05437	Adm05437 Human pro
27	363.5	8.4	1619	4	AAU00987	Aau00987 Drosophil
28	363.5	8.4	1954	8	ABM83918	Abm83918 Human dia
29	363.5	8.4	1955	8	ABM83917	Abm83917 Human dia
30	363.5	8.4	3897	8	ABM83919	Abm83919 Human dia
31	363.5	8.4	3924	8	ABM83916	Abm83916 Human dia
32	363.5	8.4	3955	8	ABM83915	Abm83915 Human dia
33	363.5	8.4	3985	8	ABM83914	Abm83914 Human dia
34	363.5	8.4	4016	8	ABM83912	Abm83912 Human dia
35	363.5	8.4	4072	8	ABM83913	Abm83913 Human dia
36	363.5	8.4	4082	8	ABM83911	Abm83911 Human dia
37	362.5	8.4	1003	8	ADS10956	Ads10956 Human the
38	358	8.3	1076	8	ADQ66099	Adq66099 Novel hum
39	355.5	8.2	843	7	ADD27861	Add27861 Rat ankyr
40	351	8.1	1709	4	AAU00971	Aau00971 Caenorhab
41	350	8.1	1088	4	AAB62429	Aab62429 Human kid
42	350	8.1	1094	6	AAE33688	Aae33688 Human str
43	350	8.1	1097	4	ABG21941	Abg21941 Novel hum
44	349.5	8.1	1453	8	ABB65398	Abb65398 Drosophil
45	348.5	8.1	1118	8	ADS10993	Ads10993 Human the

ALIGNMENTS

RESULT 1
ADO15511
ID   ADO15511 standard; protein; 831 AA.
XX
AC   ADO15511;
XX
DT   12-AUG-2004 (first entry)
XX
DE   Immunoreactive surface protein p156, SEQ ID 2.
XX
KW   Vaccine; immunoreactive; surface glycoprotein; p153; p156;
KW   canine ehrlichiosis; human ehrlichiosis.
XX
OS   Ehrlichia chaffeensis.
XX
PN   WO2004042037-A1.
XX
PD   21-MAY-2004.
XX
PF   04-NOV-2003; 2003WO-US034916.
XX
PR   04-NOV-2002; 2002US-0423573P.
XX
PA   (RBRE-) RES DEV FOUND.
XX
PI   Mcbride JW, Walker DH;
XX
DR   WPI; 2004-419705/39.
XX
PT   New DNA encoding Ehrlichia species immunoreactive surface protein p153 or
PT   protein p156, useful as vaccines against canine and human ehrlichioses or
PT   for diagnosing ehrlichioses.
XX
PS   Claim 6; SEQ ID NO 2; 41pp; English.
XX
CC   The present invention relates to Ehrlichia canis immunoreactive surface
CC   glycoprotein p153 (ADO15510) and Ehrlichia chaffeensis immunoreactive
CC   surface glycoprotein p156 (ADO15511). The proteins are useful as subunit
CC   vaccines against canine and human ehrlichioses. They are also useful for
CC   serologic and molecular diagnostics of E. canis or E. chaffeensis.
XX
SQ   Sequence 831 AA;

Query Match   100.0%; Score 4310; DB 8; Length 831;
Best Local Similarity   100.0%; Pred. No. 0;
Matches 831; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 PSGDIQDSQQDQQEQDQQQGAVGNSPIERERVAAPESEDLYTVIIPKGKRTAAPI 60

```
PN   WO200182862-A2.
XX
PD   08-NOV-2001.
XX
PF   27-APR-2001; 2001WO-US013446.
XX
PR   28-APR-2000; 2000US-00561322.
XX
PA   (RERE-) RES DEV FOUND.
XX
PI   Walker DH, Mcbride JW;
XX
DR   WPI; 2002-049309/06.
DR   N-PSDB; ADG31915.
XX
PT   New isolated and purified Ehrlichia canis immunoreactive surface protein,
PT   P43, for use as an antigen in immunodiagnosis of canine ehrlichiosis, and
PT   for preparation of a vaccine against canine ehrlichiosis.
XX
PS   Claim 10; SEQ ID NO 2; 60pp; English.
XX
CC   This invention relates to a novel use for the immunoreactive surface
CC   protein P43 specific to Ehrlichia canis (E. canis), which can be used as
CC   an antigen to immunodiagnose canine ehrlichiosis. This is a potentially
CC   fatal tick borne disease of dogs that is also known as canine monocytic
CC   ehrlichiosis, and which causes bone marrow hypoplasia as well as
CC   haematologic abnormalities such as leukopenia and thrombocytopenia. The
CC   present invention describes the serodiagnosis of canine ehrlichiosis
CC   performed by indirect fluorescent-antibody (IFA) and using recombinant E.
CC   canis P43 proteins. Accordingly, compositions derived from recombinant
CC   P43 proteins have acaricide activity and can be used in the development
CC   of a vaccine against canine monocytic ehrlichiosis. This polypeptide
CC   sequence is the E. canis P43 protein of the invention.
XX
SQ   Sequence 390 AA;

Query Match          45.5%;       Score 1963;   DB 5;   Length 390;
Best Local Similarity   99.7%;    Pred. No. 5.5e-145;
Matches  373;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1 PSGDIQDSQQDQEQDQQQGAVGGAVGNSPIERERVAAPESEDLYTVIIPKGKRTAAPI   60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16 PSGDIQDSQQDQEQDQQQGAVGGAVGNSPIERERVAAPESEDLYTVIIPKGKRTAAPI   75

Qy   61 LERKSPTPEPKVEDDEDLPPTLPPRTFSGEGYDDVGVSMPTVSRGIYQPPIVQDSNLYSS 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 LERKSPTPEPKVEDDEDLPPTLPPRTFSGEGYDDVGVSMPTVSRGIYQPPIVQDSNLYSS 135

Qy  121 IGGVPQEAQYDAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDVRNAVLGNKEIKEE 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136 IGGVPQEAQYDAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDVRNAVLGNKEIKEE 195

Qy  181 WLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVKG 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  196 WLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVKG 255

Qy  241 AKNVWTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGD 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  256 AKNVWTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGD 315

Qy  301 KLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPL 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  316 KLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPL 375

Qy  361 HHAVEHLPPVILKG 374
        |||||||| |||||
Db  376 HHAVEHLPPVNLKG 389


RESULT 3
ADO15510
ID   ADO15510 standard; protein; 831 AA.
XX
```

```
Db
Qy    1 PSGDIQDSQQDQEQDQQQGAVGGAVGNSPIERERVAAPESEDLYTVIIPKGKRTAAPI   60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db
Qy   61 LERKSPTPEPKVEDDEDLPPTLPPRTFSGEGYDDVGVSMPTVSRGIYQPPIVQDSNLYSS 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LERKSPTPEPKVEDDEDLPPTLPPRTFSGEGYDDVGVSMPTVSRGIYQPPIVQDSNLYSS 120

Qy  121 IGGVPQEAQYDAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDVRNAVLGNKEIKEE 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 IGGVPQEAQYDAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDVRNAVLGNKEIKEE 180

Qy  181 WLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVKG 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 WLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVKG 240

Qy  241 AKNVWTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGD 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 AKNVWTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGD 300

Qy  301 KLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPL 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 KLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPL 360

Qy  361 HHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAHYAVKNKNADLTLFNMLKAS 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 HHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAHYAVKNKNADLTLFNMLKAS 420

Qy  421 GADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETV 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETV 480

Qy  481 LAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKSLLMEDYK 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 LAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKSLLMEDYK 540

Qy  541 GFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATTRAALMQDRSGGELVNLGDFESENILGS 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 GFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATTRAALMQDRSGGELVNLGDFESENILGS 600

Qy  601 PNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDIINFVGDSLHLPSERGYNAMQVAALFG 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 PNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDIINFVGDSLHLPSERGYNAMQVAALFG 660

Qy  661 DKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDKN 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 DKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDKN 720

Qy  721 TVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGGDPKIVSSLLKAGAVVN 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 TVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGGDPKIVSSLLKAGAVVN 780

Qy  781 RLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADINLGDHNILFDQCL 831
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 RLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADINLGDHNILFDQCL 831


RESULT 2
ADG31916
ID   ADG31916 standard; protein; 390 AA.
XX
AC   ADG31916;
XX
DT   26-FEB-2004 (first entry)
XX
DE   E_canis P43 immunoreactive surface protein SeqID 2.
XX
KW   immunoreactive surface protein; P43; antigen; tick borne disease;
KW   canine monocytic ehrlichiosis; bone marrow hypoplasia;
KW   haematologic abnormality; leukopenia; thrombocytopenia; serodiagnosis;
KW   indirect fluorescent-antibody; IFA; acaricide; vaccine.
XX
OS   Ehrlichia canis.
XX
```

```
AC   AO015510; ;
XX
DT   12-AUG-2004   (first entry)
XX
DE   Immunoreactive surface protein p153, SEQ ID 1.
XX
KW   Vaccine; immunoreactive; surface glycoprotein; p153; p156;
KW   canine ehrlichiosis; human ehrlichiosis.
XX
OS   Ehrlichia canis.
XX
PN   WO200404042037-A1.
XX
PD   21-MAY-2004.
XX
PF   04-NOV-2003; 2003WO-US034916.
XX
PR   04-NOV-2002; 2002US-0423573P.
XX
PA   (RERE-) RES DEV FOUND.
XX
PI   Mcbride JW, Walker DH;
XX
DR   WPI; 2004-419705/39.
XX
PT   New DNA encoding Ehrlichia species immunoreactive surface protein p153 or
PT   protein p156, useful as vaccines against canine and human ehrlichioses or
PT   for diagnosing ehrlichioses.
XX
PS   Claim 12; SEQ ID NO 1; 41pp; English.
XX
CC   The present invention relates to Ehrlichia canis immunoreactive surface
CC   glycoprotein p153 (AO015510) and Ehrlichia chaffeensis immunoreactive
CC   surface glycoprotein p156 (AO015511). The proteins are useful as subunit
CC   vaccines against canine and human ehrlichioses. They are also useful for
CC   serologic and molecular diagnostics of E. canis or E. chaffeensis.
XX
SQ   Sequence 831 AA;
```

```
Query Match             30.2%;   Score 1300.5;  DB 8;   Length 831;
Best Local Similarity   35.7%;   Pred. No. 1.6e-92;
Matches  297;  Conservative 156;  Mismatches 265;  Indels 113;  Gaps  16;

Qy     39 APESEDLYTVIIPKGKRTA-APILERKSPTTEPKVEDDEDLPPTLPPRTFSGE-------- 90
             :::    ::::: : :: : :: :::::::::  :  ::     : ::    
Db     11 AVEEDPLYAVVPLPKGQRPAPTQVLEE----DPSVEEEEIAPPLPPRNNVGEVEPQEDP 65

Qy     91 -------GY-----DDVGVSMPTVSR------GIYQPPIVQDS------NLY---SSIGGVP 125
                   ::      ::  :    :         :  :   :      :        :::  :
Db     66 IYQGIPGHQEEMEEDPYASLDQVSQGAGADGIQENVPQEAGEELEEDIYQDPADFQGLG 125

Qy    126 QEA-QYDAAARAG--GPRKFLYGPYTFSNGQEIMDFEFDTPWPDVRNAVLGNKEIKEEW 181
             :::   :   :::      ::: ::   :   :::    ::::::: ::  ::::  :
Db    126 QGGQOLDQAGYQGPSIGDRQLVNGPYGFNDGSYAMEFD-DVMWEGVRDAVIHDEEIDPKF 184

Qy    182 LTTSGPVRDIADRIVASKGDLSEBDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVKGA 241
             : : :::: :: :: ::::  :::  :::  :  ::     ::::  :  :::::   ::
Db    185 LVTDGLMRHICDKIVQSEGNLPEPDLEEIVSILKNDKEGISELINEPVQVDIPNNPVREG 244

Qy    242 KNVMTLMHLVACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGDK 301
             :||||| :|  :     :|||| | :    ::::  :||||||||||||||||| ||
Db    245 RNVMTLLHLHAYAYNVDPRIINAIESVENSFGESGLDGYNIQDADGNLPLHHAAKNCNGQV 304

Qy    302 LKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVD-GDGKNPL 360
             : :|:|||| |   ::| ||||||  ::::::|  : ::::::::|:  : :  |: ::
Db    305 LDNCISKTNSNIINIRNFGNQSPLHVMVQNPGCSIGNIQVANECGMDFNLIDHPTGRMPI 364

Qy    361 HHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYPGNTIAHYAVKNADLTLFNMLKAS 420
             ||||    ::  :: ::::: : ::   ::      ::
Db    365 HYAAE-------------AASSE---------------VLSYVIRNTKAE--------- 386

Qy    421 GADLNRVNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETV 480
             ::|||    : :::  :|:|  |||  :  :|:|| : :: :||:|| |||||  |:   :
```

```
Db    387 ------------------------------------SPQASA-----------------VNTQDVNGRTPLHCAAISGNSKGL 416
XX
Qy    481 LAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKSLLMEDYK 540
             |||||:|||||||:: ::  ::|:||||||||||| || :||| ::  ||||| |::|||
Db    417 SVMLLQNGVDCAVRDKNYSTPLHYAVAGNDIKSIKNLCSVKGRVQGVKSSAASLLCEDLQ 476
XX
Qy    541 GFTPLHFVAGGSRDTFPRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGS 600
             | ||||| |:  ::|  :|::: ::|||||||||:|||:|:::: :||::|  :::|||
Db    477 GDTPLHIACKVEGTKAFETVRQSIKKHHGKQVLQELLIREGSGPRL-NVSGFGSQSILSG 535
XX
Qy    601 PNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFG 660
             |:||||| ||::| | |||||:|||||  :|||||:|||| ||||:||||: ||::||||
Db    536 VSGDLYGYLNSQNFPTSPVHAAVKANNLQLLNLFLKKSPDILRQSSPNGFNPVHMAALFA 595
XX
Qy    661 DKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDKN 720
             |::|:|:|: :::|||| |:::|::::::|:| ||| | ::| ||::::|| ::  ||::|
Db    596 DVKTVKLIIENASGEEVNAQSDSTLTPLHLACIRGDGSIIKRMVEHESVNVNQTMGPDQN 655
XX
Qy    721 TVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHIAVEGGDPKIVSSLLKAGAVVN 780
             |:|  :  ::|::|:: |::||| | ||  ||:|||||:|:|:|:|| | | || |||||
Db    656 TVLQYAINRGNHSLIKRLLSHPSIDLNVRNADGKTSAHSAMEKGDLKTVKALCNAGADVN 715
XX
Qy    781 RLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADINLDGDHNILFDQCL 831
             : ::||||: ::| : |::   |||:| ||::: ||| | ||: : : |||
Db    716 TVDNNGRSVISSAIYSGQNEKKLVPIVKLLNSGAKIGSQEDKNILLQKCI 766
```

```
RESULT 4
AAU28174
ID   AAU28174 standard; protein; 1053 AA.
XX
AC   AAU28174;
XX
DT   18-DEC-2001   (first entry)
XX
DE   Novel human secretory protein, Seq ID No 343.
XX
KW   Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
KW   ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
KW   transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
KW   amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
KW   ulcer; osteoporosis; bone degenerative disease; periodontal disease;
KW   gut protection; lung; liver fibrosis; immune deficiency; infection;
KW   severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
KW   multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
KW   fertility; analgesic; pain; antigen.
XX
OS   Homo sapiens.
XX
PN   WO200166689-A2.
XX
PD   13-SEP-2001.
XX
PF   05-MAR-2001; 2001WO-US004942.
XX
PR   07-MAR-2000; 2000US-00519705.
PR   19-MAY-2000; 2000US-00574454.
PR   17-JUN-2000; 2000US-00596193.
PR   14-JUL-2000; 2000US-00616847.
PR   19-SEP-2000; 2000US-00665363.
PR   20-OCT-2000; 2000US-00693267.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;
PI   Zhao QA, Yang Y, Drmanac RT, Zhang Y, Chen R, Xue AJ, Wang J;
XX
DR   WPI; 2001-589934/66.
DR   N-PSDB; AAS45074.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis and treatment of
PT   cancer, neurological, inflammatory, and autoimmune disorders.
```

```
XX
PS   Example 5; SEQ ID NO 343; 107pp; English.
XX
CC   The invention relates to novel isolated human secreted polypeptides (I)
CC   and polynucleotides (II). (I) and (II) are useful for treating
CC   inflammatory conditions such as arthritis, nephritis, Crohn's disease,
CC   ischaemia-reperfusion injury, shock, sepsis, immune responses, and is
CC   involved in increasing haematopoiesis, stem cell survival, bone growth
CC   and remodeling. (I), (II) and modulators of (II) are useful for
CC   prophylaxis or treatment of one or more cancers. (II) is also useful for
CC   creating transgenic animals useful for studying the in vivo activities of
CC   the polypeptide as well as for studying modulators of the polypeptides.
CC   (I) induces the proliferation of neural cells and regeneration of nerve
CC   and brain tissue and is useful for the treatment of central and
CC   peripheral nervous system diseases and neuropathies, such as Alzheimer's,
CC   Parkinson's disease, Huntington's disease, and amyotrophic lateral
CC   sclerosis. In addition, (I) is involved in chemotactic or chemokinetic
CC   activity, regulation of haematopoiesis and is useful for treating myeloid
CC   or lymphoid cell disorders. Platelet disorders such as thrombocytopenia
CC   and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC   tissue growth, and in tissue repair, healing of burns, incisions, ulcers,
CC   for treating osteoporosis, osteoarthritis, bone degenerative disorders,
CC   or periodontal disease. Furthermore, (I) is also useful for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues, various immune deficiencies and
CC   disorders including severe combined immunodeficiency (SCID), bacterial or
CC   fungal infections, autoimmune disorders e.g. multiple sclerosis,
CC   rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
CC   reactions and conditions, such as asthma or other respiratory problems.
CC   In addition, (I) affects biorhythms or circadian cycles of rhythms,
CC   fertility, metabolism, catabolism, anabolism, storage or elimination of
CC   dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
CC   analgesic effects or other pain reducing effects, immunoglobulin like
CC   activity and can act as an antigen in a vaccine composition to raise an
CC   immune response. AAU28020-AAU28395 represent novel human secreted protein
CC   amino acid sequences of the invention
XX
SQ   Sequence 1053 AA;

Query Match      9.8%;    Score 421;   DB 4;   Length 1053;
Best Local Similarity   25.2%;   Pred. No. 2.1e-23;
Matches  189;  Conservative 101;  Mismatches 269;  Indels 190;  Gaps  29;

Qy  170 AVLGNKEIKE------------EWLTTSGPVRDIADRIVASKGDLSEDQVEILDIIF 215
            :|: | ::|::            |||:    ||::: : :::||
Db   49 AYLGDAEIIBLLILSGARVNAKDSKWLT---PLH---RAVASCS------EEAVQVLL 94

Qy  216 MNESEIAEGISN--PLHADVDNNPVKGAKNVMTLMHLYACDVDPRIVKALGEVENDEG 272
            ::|:|  ||    |||:| :|: :|:::|  |::|
Db   95 KHSADVNARDKNWQTPLHIAAANKAVKCAEALVPLLSNV------------------ 133

Qy  273 DLGANAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKP 332
            |||||| |||   |||:|||
Db  134 ------NVSDRAGRTALHHAAFSGHGEMVKLLLSRGAN--INAFDKKDRRAIHWAAYMG 184

Qy  333 DCSVLDIEEFTSRNLDFGLVDGDGKGNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDP 392
            : :                  ::    |||:|: :            :|
Db  185 HIEVVKL--LVSHGAEVTCKDKKSYTPLHAAAS------SGMISVVKYLLDLGVDMNEP 235

Qy  393 DYFGNTIAHYAVKNKNADLTLPNMLKASGADLNVRNVVGRAPIH-VASSNGKANAVSGLV 451
            : |||| :| |:|      |     ::   |||:   |  :|:|   ||:||
Db  236 NAYGNTPLHVACYN-GQDVVV-NELIDCGAIVNQKNEKGFTPLHFAAASTHGALCLELLIV 293

Qy  452 SCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDI 511
             |||:||: | ||||||| |      | :| ::|:|||:||:  |||||:::|||:||:
Db  294 GNGADVNMSKDGKTPLHMTALHGRFSRSQTII-QSGAVIDCEDKNGNTPLHIAARYGHE 352

Qy  512 GVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVACGGSGSRDTFRVVRKNYEKCHDLA 571
            |:| ||   :  : ||||| | |:        |||| |  || | : : : :   ::  | |
Db  353 LLINTLITS----GADTAKRGI----HGMFPLHLAALLSGFSD---CCRKLLSSGFDID 399

Qy  572 TI----RAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSN 627
            |:     |      ||   ||| |        |  |    :|  ||  | |  |  ::::
Db  400 TPDDFGRTCLHAAAAGGNL------ECLNLLLNTGADF---NKKDKFGRSPLHYAAANCN 450

Qy  628 HNVMKDILN--FVGD------------SLH------------LPSERGYNAMQ 654
            :::::|:|:    ||              ||            |::| |||| |
Db  451 YQCLFALVGSGASVNDLDERGCTPLHYAATSDTGKCLEYLLRNDANPGIRDKQGYNAVH 510

Qy  655 VAALFGDKEAVKMLAKSAKPSDLNFKTS------------ATPTPLNLACLRGDNEVVRGL 703
            ::|  |:  |:  |:                          :||||| |:   |:  : |
Db  511 YSAAYGHRLCLQLIA-SETPLDVLMETSGTDMLSDSDNRATISPLHLAAYHGHHQALEVL 569

Qy  704 VGQHGIDINQRMGSD------------------KNTVLHYAISKG 730
            |  : || :|                       :::|:|||:|:|
Db  570 V-QSLLDLDVRNSSGRTPLDLAAFKGHVECVDVLINQGASILVKDYILKRTPIHAAATNG 628

Qy  731 DSFLVQKILAH------TGVDVNCENNLGQTPLHIAVEGGDPKIVSLLKAGAVVNRLDDNG 786
            :|     :| :        |:||    ||||||::|::|    :|:||: ||:||  | :|
Db  629 HSECLRLLIGNAEFQNAVDIQDGN--GQTPLMLSVLNGHTDCCVYSLLNKGANVDAKDKWG 686

Qy  787 RSVLSSAIVPGRKEKGVLGIVNKLLDRGA 815
            |::|  |:     :  |:  :::|||:|||
Db  687 RTALHRGAVTGHEE------CVDALLQHGA 710
```

RESULT 5
ADK65850

```
ID   ADK65850 standard; protein; 1059 AA.
XX
AC   ADK65850;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Angiogenesis-differentially expressed protein #60.
XX
KW   cytostatic; cardiant; vasotropic; antiarteriosclerotic;
KW   angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;
KW   gene expression; cancer; coronary artery disease; myocardial ischemia;
KW   coronary arteriosclerosis; forensic medicine.
XX
OS   Homo sapiens.
XX
PN   WO200306831-A2.
XX
PD   14-AUG-2003.
XX
PF   07-FEB-2003; 2003WO-US003848.
XX
PR   07-FEB-2002; 2002US-00067482.
PR   10-JUN-2002; 2002US-00164595.
PR   16-AUG-2002; 2002US-0403649P.
PR   03-JAN-2003; 2003US-0437746P.
XX
PA   (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI   Sun Z,  Li X,  Kovacs KF,  Fan W,  Jay G;
XX
DR   WPI; 2003-731502/69.
XX
PT   Determining the angiogenic index of a tissue or cell sample using
PT   expression levels of differentially expressed genes, useful for
PT   diagnosing or treating cancer, coronary artery disease, myocardial
PT   ischemia and/or arteriosclerosis.
XX
PS   Disclosure; SEQ ID NO 89; 296pp; English.
XX
CC   The invention relates to a method of determining the angiogenic index of
CC   a tissue or cell sample comprising assessing, in a sample, the expression
CC   levels of one or more differentially-expressed gene from any of 34 DNA
CC   sequences, given in the specification, where the levels are indicative of
CC   the angiogenic index. The methods and compositions of the present
CC   invention are useful for diagnosing, preventing and/or treating cancer,
CC   coronary artery disease, myocardial ischemia or coronary
CC   arteriosclerosis. They can also be used in research, drug discovery and
CC   forensic medicine involving angiogenesis. This sequence corresponds to
```

KW   inflammatory bowel disease; infectious disease; septic shock;
KW   bacterial infection; neurological disease; stroke-induced inflammation;
KW   neurodegenerative disease; cancer; antiinflammatory; antiarthritic;
KW   antirheumatic; cytostatic; antibacterial; gene therapy; human.
OS   Homo sapiens.
XX
PN   WO2004035783-A2.
XX
PD   29-APR-2004.
XX
PF   24-SEP-2003; 2003WO-EP050655.
XX
PR   26-SEP-2002; 2002EP-00021809.
PR   10-FEB-2003; 2003EP-00100274.
XX
PA   (CELL-) CELLZOME AG.
XX
PI   Bouwmeester T, Huhse B, Bauch A, Ruffner H, Bauer A, Kuester B;
PI   Superti-Furga G, Kruse U;
XX
DR   WPI; 2004-348460/32.
XX
PT   New protein complex comprising at least one first and second protein of
PT   the Tumor Necrosis Factor-alpha (TNF-alpha)-signaling pathway, useful for
PT   diagnosing or treating inflammation, neurological diseases, infectious
PT   diseases or cancer.
XX
PS   Example; SEQ ID NO 185; 1980pp; English.
XX
CC   This invention relates to novel protein complexes of the tumour necrosis
CC   factor-alpha (TNF-alpha) signalling pathway. Specifically, it refers to
CC   methods for preparing these complexes comprising at least two component
CC   proteins, as well as screening methods to identify modulators of the
CC   pathway, which include antibodies, agonists and antagonists thereof. The
CC   present invention describes a protein complex and kit that are useful for
CC   diagnosing, prognosing or treating chronic inflammatory diseases such as
CC   rheumatoid arthritis and inflammatory bowel disease; infectious diseases
CC   such as septic shock and bacterial infections; neurological diseases such
CC   as stroke-induced inflammation in neurons; neurodegenerative diseases and
CC   cancer. Accordingly, these complexes can be used for the development of
CC   pharmaceutical compositions that exhibit antiinflammatory, antiarthritic,
CC   antirheumatic, cytostatic and antibacterial activities and can be used
CC   for gene therapy purposes. In particular, the invention further provides
CC   siRNA-oligonucleotides useful for inhibiting protein expression for in
CC   vitro or cell culture assays. This polypeptide is a human protein that
CC   can be used in combination with other proteins provided in the
CC   specification to form novel complexes of the TNF-alpha signalling pathway
CC   of the invention.
XX
SQ   Sequence 1059 AA;

Query Match   9.8%;   Score 421; DB 8;   Length 1059;
Best Local Similarity 25.2%; Pred. No. 2.2e-23;
Matches 189; Conservative 101; Mismatches 269; Indels 190; Gaps   29;

Qy   170 AVLGNKEIKE-----------------EWLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIF 215
                ||:||| :|||: ||: ||
Db   55 AYLGDAEIIELLILSGARVNAKDSKWLT---PLH----RAVASCS------EEAVQVLL 100

Qy   216 MNESEIAEGISN---PLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEG 272
                ::::: ||| :|: :|::|:
Db   101 KHSADVNARDKNWQTPLHIAAANKAVKCAEALVPLLSNV----------------- 139

Qy   273 DLGANAYNVLDSEGNLPLHHAAKNCTGDKLKJCMEKTKTDFIDTANFANQSPLHIITQKP 332
                      ||: |||| ||
Db   140 ------NVSDRAGRTALHHAAFSGHGEMVKLLLSRGAN--INAFDKKDRRAIHWAAYMG 190

Qy   333 DCSVLDIEEFTSRNLDPFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDP 392
                :|: || :|: ||
Db   191 HIEVVKL--LVSHGAEVTCKDKKSYTPLHAAAS------SGMISVVKYLLDLGVDMNEP 241

Qy   393 DYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIH-VASSNGKANAVSGLV 451
                : |||| | ::: | :: |||||| :
Db   242 NAYGNTPLHVACYN-GQDVVV-NELIDCGAIVNQKNEKGFTPLHFAAASTHGALCLELLV 299

Qy   452 SCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDI 511
                |: :|| :|| || :
Db   300 GNGADVNMKSKDGKTPLHMTALHGRFSRSQTII-QSGAVIDCEDKNGNTPLHIAARYGHE 358

Qy   512 GVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFPRVVRKNYEKCHDLA 571
                ::   :|::
Db   359 LLINTLITS-----GADTAKRGI-----HGMFPLHLAALSGFSD---CCRKLLSSGFDID 405

Qy   572 TI----RAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSN 627
                |: |:   :
Db   406 TPDDFGRTCLHAAAAGNL--------ECLNLLLNTGADF---NKKDKFGRSPLHYAAANCN 456

Qy   628 HNVMKDILN---FVGD------SLH--------------------LPSERGYNAMQ 654
                 :|:   ::| :|| |
Db   457 YQCLFALVGSGASVNDLDERGCTPLHYAATSDTDGKCLEYLLRNDANPGIRDKQGYNAVH 516

Qy   655 VAALFGDKEAVKMLAKSAKPSDLNFKTS--------ATPTPLNLACLRGDNEVVRGL 703
                |::|| ::|::| :| || || | ||| ||
Db   517 YSAAYGHRLCLQLIA-SETPLDVLMETSGTDMLSDSDNRATISPLHLAAYHGHHQALEVL 575

Qy   704 VGQHGIDINQRMGSD-----------------------------KNTVLHYAISKG 730
                | |||| |: ||
Db   576 V-QSLLDLDVRNSSGRTPLDDLAAFKGHVECVDVLINQGASILVKDYILKRTPIHAAATNG 634

Qy   731 DSFLVQKILAH----TGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRLDDNG 786
                 |:| |:| |:||| || ||| :| || || ||:|
Db   635 HSECLRLLIGNAEPQNAVDIQDGN--GQTPLMLSVLNGHTDCVYSILLNKGANVDAKDKWG 692

Qy   787 RSVLSSAIVPGRKEKGVLGIVNKLLDRGA 815
                | |: |::| ||
Db   693 RTALHRGAVTGHEE------CVDALLQHGA 716

RESULT 6
ADS88330
ID   ADS88330 standard; protein; 1059 AA.
XX
AC   ADS88330;
XX
DT   18-NOV-2004 (first entry)
XX
DE   Human protein of a TNF-alpha signalling pathway protein complex Seq 185.
XX
KW   protein complex; tumour necrosis factor-alpha signalling pathway;
KW   TNF-alpha; chronic inflammatory disease; rheumatoid arthritis;

CC   one of the differentially expressed proteins of the invention.
XX
SQ   Sequence 1059 AA;

Query Match   9.8%;   Score 421; DB 7;   Length 1059;
Best Local Similarity 25.2%; Pred. No. 2.2e-23;
Matches 189; Conservative 101; Mismatches 269; Indels 190; Gaps   29;

Qy   170 AVLGNKEIKE-----------------EWLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIF 215
                ||:||| :|||: ||: ||
Db   55 AYLGDAEIIELLILSGARVNAKDSKWLT---PLH----RAVASCS------EEAVQVLL 100

Qy   216 MNESEIAEGISN---PLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEG 272
                ::::: ||| :|: :|::|:
Db   101 KHSADVNARDKNWQTPLHIAAANKAVKCAEALVPLLSNV----------------- 139

Qy   273 DLGANAYNVLDSEGNLPLHHAAKNCTGDKLKJCMEKTKTDFIDTANFANQSPLHIITQKP 332
                      ||: |||| ||
Db   140 ------NVSDRAGRTALHHAAFSGHGEMVKLLLSRGAN--INAFDKKDRRAIHWAAYMG 190

Qy   333 DCSVLDIEEFTSRNLDPFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDP 392
                :|: || :|: ||
Db   191 HIEVVKL--LVSHGAEVTCKDKKSYTPLHAAAS------SGMISVVKYLLDLGVDMNEP 241

Qy   393 DYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIH-VASSNGKANAVSGLV 451
                : |||| | ::: | :: |||||| :
Db   242 NAYGNTPLHVACYN-GQDVVV-NELIDCGAIVNQKNEKGFTPLHFAAASTHGALCLELLV 299

Qy   452 SCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDI 511
                |: :|| :|| || :
Db   300 GNGADVNMKSKDGKTPLHMTALHGRFSRSQTII-QSGAVIDCEDKNGNTPLHIAARYGHE 358

Qy   512 GVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFPRVVRKNYEKCHDLA 571
                ::   :|::
Db   359 LLINTLITS-----GADTAKRGI-----HGMFPLHLAALSGFSD---CCRKLLSSGFDID 405

Qy   572 TI----RAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSN 627
                |: |:   :
Db   406 TPDDFGRTCLHAAAAGNL--------ECLNLLLNTGADF---NKKDKFGRSPLHYAAANCN 456

Qy   628 HNVMKDILN---FVGD------SLH--------------------LPSERGYNAMQ 654
                 :|:   ::| :|| |
Db   457 YQCLFALVGSGASVNDLDERGCTPLHYAATSDTDGKCLEYLLRNDANPGIRDKQGYNAVH 516

Qy   655 VAALFGDKEAVKMLAKSAKPSDLNFKTS--------ATPTPLNLACLRGDNEVVRGL 703
                |::|| ::|::| :| || || | ||| ||
Db   517 YSAAYGHRLCLQLIA-SETPLDVLMETSGTDMLSDSDNRATISPLHLAAYHGHHQALEVL 575

Qy   704 VGQHGIDINQRMGSD-----------------------------KNTVLHYAISKG 730
                | |||| |: ||
Db   576 V-QSLLDLDVRNSSGRTPLDDLAAFKGHVECVDVLINQGASILVKDYILKRTPIHAAATNG 634

Qy   731 DSFLVQKILAH----TGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRLDDNG 786
                 |:| |:| |:||| || ||| :| || || ||:|
Db   635 HSECLRLLIGNAEPQNAVDIQDGN--GQTPLMLSVLNGHTDCVYSILLNKGANVDAKDKWG 692

Qy   787 RSVLSSAIVPGRKEKGVLGIVNKLLDRGA 815
                | |: |::| ||
Db   693 RTALHRGAVTGHEE------CVDALLQHGA 716

```
Db            : :||||  |||    |:.|   |||.| :||   |  ||.|  |:|   ||:|
        242 NAYGNTPLHVACYN-GQDVVV-NELIDCGAIVNQKNEKGFTPLHFAAASTHGALCLELLV 299

Qy     452 SCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDI 511
Db            : :||||  .:|.||||:         ::  ||  ||.:|||: :||:| |||:|
        300 GNGADVNMKSKDGKTPLHMTALHGRFSRSQTII-QSGAVIDCEDKNGNTPLHIAARYGHE 358

Qy     512 GVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLA 571
Db            :.|:   ||:        :|.:          |:
        359 LLINTLITS----GADTAKRGI-----HGMFPLHLAALSGFSD---CCRKLLSSGFDID 405

Qy     572 TI----RAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSN 627
Db            |   ||   |||    ||                           : :|
        406 TPDDFGRTCLHAAAAGGNL------ECLNLLLNTGADF--NKKDKFGRSPLHYAAANCN 456

Qy     628 HNVMKDILN---FVGD------SLH------------------------LPSERGYNAMQ 654
Db            : .:                                             :||||:
        457 YQCLFALVGSGASVNDLDERGCTPLHYAATSDTDGKCLEYLLRNDANPGIRDKQGYNAVH 516

Qy     655 VAALFGDKEAVKMLAKSAKPSDLNPKTS---------ATPTPLNLACLRGDNEVVRGL 703
Db            :|||  :::|:  .:|                       :||||.||||:||||
        517 YSAAYGHRLCLQLIA-SETPLDVLMETSGTDMLSDSDNRATISPLHLAAYHGHHQALEVL 575

Qy     704 VGQHGIDINQRMGSD---------------------KNTVLHYAISKG 730
Db            |   :| |:                         |  :||| :||.
        576 V-QSLLDLDVRNSSGRTPLDLAAFKGHVECVDVLINQGASILVKDYILKRTPIHAAATNG 634

Qy     731 DSFLVQKILAH----TGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVNRLDDNG 786
Db            :.: ::|     ||             |||||||:  ||||||.|||||||:||::||
        635 HSECLRLLIGNAEPQNAVDIQDGN--GQTPLMLSSVLNGHTDCVYSLLNKGANVDAKDKWG 692

Qy     787 RSVLSSAIVPGRKEKGVLGIVNKLLDRGA 815
Db            |.||  ||      :|:||||||||:.|
        693 RTALHRGAVTGHEE-----CVDALLQHGA 716
```

RESULT 7
AAU20496
ID AAU20496   standard; protein; 705 AA.
XX
AC AAU20496;
XX
DT 06-DEC-2001 (first entry)
XX
DE Human secreted protein, Seq ID No 488.
XX
KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
KW Gaucher's disease; neurological disease; cerebrovascular disorder;
KW thrombosis; wound healing.
XX
OS Homo sapiens.
XX
PN WO200155326-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001347.
XX

PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0233398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0230064P.
PR 14-SEP-2000; 2000US-0230065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.

```
PR   20-OCT-2000;  2000US-0241808P.
PR   20-OCT-2000;  2000US-0241809P.
PR   20-OCT-2000;  2000US-0241826P.
PR   01-NOV-2000;  2000US-0244617P.
PR   08-NOV-2000;  2000US-0246474P.
PR   08-NOV-2000;  2000US-0246475P.
PR   08-NOV-2000;  2000US-0246476P.
PR   08-NOV-2000;  2000US-0246477P.
PR   08-NOV-2000;  2000US-0246478P.
PR   08-NOV-2000;  2000US-0245523P.
PR   08-NOV-2000;  2000US-0245524P.
PR   08-NOV-2000;  2000US-0246525P.
PR   08-NOV-2000;  2000US-0245526P.
PR   08-NOV-2000;  2000US-0245527P.
PR   08-NOV-2000;  2000US-0246528P.
PR   08-NOV-2000;  2000US-0246532P.
PR   08-NOV-2000;  2000US-0246609P.
PR   08-NOV-2000;  2000US-0246610P.
PR   08-NOV-2000;  2000US-0246611P.
PR   08-NOV-2000;  2000US-0246613P.
PR   17-NOV-2000;  2000US-0249207P.
PR   17-NOV-2000;  2000US-0249208P.
PR   17-NOV-2000;  2000US-0249209P.
PR   17-NOV-2000;  2000US-0249210P.
PR   17-NOV-2000;  2000US-0249211P.
PR   17-NOV-2000;  2000US-0249212P.
PR   17-NOV-2000;  2000US-0249213P.
PR   17-NOV-2000;  2000US-0249214P.
PR   17-NOV-2000;  2000US-0249215P.
PR   17-NOV-2000;  2000US-0249216P.
PR   17-NOV-2000;  2000US-0249217P.
PR   17-NOV-2000;  2000US-0249218P.
PR   17-NOV-2000;  2000US-0249244P.
PR   17-NOV-2000;  2000US-0249245P.
PR   17-NOV-2000;  2000US-0249264P.
PR   17-NOV-2000;  2000US-0249265P.
PR   17-NOV-2000;  2000US-0249297P.
PR   17-NOV-2000;  2000US-0249299P.
PR   17-NOV-2000;  2000US-0249300P.
PR   01-DEC-2000;  2000US-0250160P.
PR   01-DEC-2000;  2000US-0250391P.
PR   05-DEC-2000;  2000US-0251030P.
PR   05-DEC-2000;  2000US-0251869P.
PR   05-DEC-2000;  2000US-0251988P.
PR   06-DEC-2000;  2000US-0256719P.
PR   06-DEC-2000;  2000US-0251479P.
PR   08-DEC-2000;  2000US-0251856P.
PR   08-DEC-2000;  2000US-0251868P.
PR   08-DEC-2000;  2000US-0251869P.
PR   08-DEC-2000;  2000US-0251989P.
PR   08-DEC-2000;  2000US-0251990P.
PR   11-DEC-2000;  2000US-0254097P.
PR   05-JAN-2001;  2001US-0259678P.
```

XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-451931/48.
DR   N-PSDB; AAS33205.
XX
PT   New nucleic acids and polypeptides, useful for diagnosing, preventing or
PT   treating medical conditions.
XX
PS   Claim 11; SEQ ID NO 488; 753pp; English.
XX
CC   The invention relates to novel isolated nucleic acid molecules (I)
CC   encoding human secreted proteins (II). (I) and (II) are used to prevent,
CC   treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
CC   goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
CC   the prevention, treatment and diagnosis of diseases associated with
CC   inappropriate expression of secreted proteins. (I) and complementary
CC   sequences may also be used as DNA probes in diagnostic assays (e.g.
CC   polymerase chain reactions (PCR)) to detect and quantitate the presence
CC   of similar nucleic acid sequences in samples, and so which patients may
CC   be in need of restorative therapy. (II) may also be used as antigens in
CC   the production of antibodies and in assays to identify modulators
CC   (agonists and antagonists) of the expression and activity of the secreted
CC   proteins. The anti-(II) antibodies and antagonists may also be used to
CC   down regulate expression and activity of (II). The anti-(II) antibodies
CC   may also be used as diagnostic agents for detecting the presence of (II)
CC   in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The
CC   disorders include for example: immune/autoimmune diseases (e.g. HIV
CC   (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC   and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC   melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC   Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC   Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
CC   cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and
CC   thrombosis), infections caused by bacteria, viruses and fungi and ocular
CC   disorders (e.g. corneal infections). (I) and (II), agonists, antagonists
CC   and antibodies can also be used to promote wound healing, maintain organs
CC   before transplantation, and support cell culture of primary tissues.

```
Query Match        9.6%;  Score 413;  DB 4;  Length 705;
Best Local Similarity  25.2%;  Pred. No. 4.9e-23;
Matches 189;  Conservative 99;  Mismatches 271;  Indels 190;  Gaps 29;

Qy  170 AVLGNKEIKE----------------EWLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIF 215
            ||:.|||         ||        |||  ::|:
Db   26 AYLGDAEIIELLILSGARVNAKDSKWLT--PLH---RAVASCS-------EEAVQVLL 71

Qy  216 MNESEIAEGISN---PLHADVDNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEG 272
            : .::          |||         |:  :  : :||||:
Db   72 KHSADVNARDKNWQTPLHIAAANKAVKCAEALVPLLSNV----------------- 110

Qy  273 DLGANAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFQNQSPLHIITQKP 332
            :||||     :|:.|||            :||||
Db  111 ------NVSDRAGRTALHHAAFSGHGEMVKLLSRGAN--INAFDKKDRRAIHWAAYMG 161

Qy  333 DCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEBHLPPVILKGVMDHVKNSSEFQDLVNDP 392
            :||:                     :.::         |: :||.:
Db  162 HIEVVKL--LVSHGAEVTCKDRKSYTPLHAAAS---------SGMISVVKYLLDLGVDMNEP 212

Qy  393 DYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIH-VASSNGKANAVSGLV 451
            ||||                          :||||
Db  213 NAYGNTPLHVACYN-GQDVVV-NELIDCGAIVNQKNEKGFTPLHFAAASTHGALCLELLV 270

Qy  452 SCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDI 511
            ||||                          :.:|||
Db  271 GNGADVNMKSDGKTPLHMTXXHGRFSRSQTII-QSGAVIDCEDKNGNTPLHIAARYGHE 329

Qy  512 GVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGGSRDTFRVVRKNYEKCHDLA 571
            ||||                          |||
Db  330 LLINTLJTS-----GADTAKRGI------HGMPFLHLAALSGFSD---CCRKLLSSGPFID 376

Qy  572 TI----RAALMQDRSGGELVNLGDFSENILGSPNAKFPLQHIQSANFGFSPARRGIVSSN 627
            |||                            :.:.:::
Db  377 TPDDFGRTWSTCSCSGGNL-------ECLNLLNTGADF---NKKDKFGRSPLHYAAANCN 427

Qy  628 HNVMKDILN---FVGD--------SLH------------------LPSERGYNAMQ 654
            : :.:|||                :.|||||:
Db  428 YQCLFALVGSGASVNDLDERGCTLPHYAATSXTDGKCLEYLLRNDANPGIRDKQGYNAVH 487

Qy  655 VAALFGDKEAVKMLAKSAKPSDLNFKTS---------ATPTPLNLACLRGDNEVVRGL 703
            ||||:                          |:.:|::
Db  488 YSAAYGHRLCXQLIA-SETPLDVLMETSGTDMLSDSDNRATISPLHLAAYHGHHQALRVL 546

Qy  704 VGQHGIDINQRMGSD------------------------------KNTVLHYAISKG 730
            |||:                             ::.::
Db  547 V-QSLLDLDVRNSSGRTPLDLAAFKGHVECVDVLINQGASIIVKDYILKRTPIHAAAATNG 605

Qy  731 DSFLVQKILAH----TGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRLDDNG 786
            |||||                         :.:::||||
Db  606 HSECLRLLIGNAEPQNAVDIQDGN--GQTPLMLSVLNGHTDCVYSLLNKGANVDAKDKWG 663

Qy  787 RSVLSSAIVPGRKEKGVLGIVNKLLDRGA 815
```

```
Db   461 ------------------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP 506
              |:.|  || |      |:.||.||
Db   664 RTALHRGAVTGHEE-------CVDALLQHGA 687

RESULT 8
ADR90358
ID   ADR90358 standard; protein; 1880 AA.
XX
AC   ADR90358;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Full length human ankyrin, seq id 2.
XX
KW   Antiparasitic; gene therapy; falcipain-2; inhibitor; ankyrin;
KW   erythrocyte protein 4.1; haemoglobin; malaria; Ank P1; human.
XX
OS   Homo sapiens.
XX
PN   WO2004052914-A2.
XX
PD   24-JUN-2004.
XX
PF   05-DEC-2003; 2003WO-US038835.
XX
PR   06-DEC-2002; 2002US-0431294P.
XX
PA   (SELI-) ST ELIZABETH'S MEDICAL CENT BOSTON INC.
XX
PI   Hanspal M, Chishti AH;
XX
DR   WPI; 2004-468808/44.
DR   N-PSDB; ADR90359.
XX
PT   New falcipain-2 inhibitor peptide, useful for preparing a pharmaceutical
PT   composition for treating malaria infection caused by Plasmodium
PT   falciparum.
XX
PS   Example 6; SEQ ID NO 2; 72pp; English.
XX
CC   The invention relates to an isolated falcipain-2 inhibitor peptide.
CC   Peptides of the invention are derived from ankyrin and erythrocyte
CC   protein 4.1, and selectively inhibit falcipain-2-mediated cleavage of
CC   human erythrocyte protein 4.1, erythrocyte membrane ankyrin and
CC   haemoglobin. Further disclosed is a method for treating malaria infection
CC   in a subject suspected of having malaria. The falcipain-2 inhibitor
CC   peptide is useful for preparing a composition for treating malaria
CC   infection. The current sequence is full length human ankyrin.
XX
SQ   Sequence 1880 AA;

Query Match             9.1%;  Score 391.5; DB 8; Length 1880;
Best Local Similarity   26.0%;  Pred. No. 1.1e-20;
Matches  174;     Conservative 100;  Mismatches 246;   Indels 149;   Gaps  31;

Qy   194 RIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVK--------GA---- 241
             | : ||||   : : |||     |||    :  ||
Db   256 RLLLDRGAQETKTKDEL------------------TPLHCAARNGHVRISEILLDHGAPIQA 300

Qy   242 --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT 298
           | :.|  |||||    : |||:|             :  |||  |||||||||
Db   301 KTKNGLSPIHMAAQGDHLD--CVRLLQYDAEIDDI------TLD---HLTPLHVAAHCG 349

Qy   299 GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK 357
             :: .:||                     ||| :     |:||    ::
Db   350 HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL 405

Qy   358 NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT-IAHYAVKNK 407
           ||||||   ||||||    :||     ||||:|:   : |||    ||    :|  | ||
Db   406 TPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK 460

Qy   408 NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP 467
           |||  :  || ||  | :  : |||:  :|:| |
```

```
Db   461 ------------------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP 506
             :|| |||:  : |    :    ||||
Qy   468 LHIAVEGGSMETVLAVLNQRGADVSQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK 525
             |:|||  | || ||| | | |
Db   507 LHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA 565
             ||| :: |   :|||||::  |
Qy   526 GEDTVAKSILMEDYKGFTPLHFVAGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE 585
             |:      :            ::
Db   566 GKN------------GLTPLHVAVHHNNLDIVKL------------------------- 587

Qy   586 LVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDIINFVGDSLHLP 645
             ||||:          |||    ||                    |||||  :||||
Db   588 LLPRG--------GSP-------HSPAWN-GYTPLHIAAKQNQVEVARSLLQY-GGSANAE 631

Qy   646 SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG 705
             :||   : |||  ||||  :      :  :    |||||  || : :|||||  |
Db   632 SVQGVTPLHLAAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHLVAQEGHGHVPVADVLI- 688

Qy   706 QHG---IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG 763
             :|:    | |      :  |||  || |||||:|:|| ||  |||  || ||||||||
Db   689 KHGVMVDATTRMG----YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ 744

Qy   764 GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL---LDRGADINLD 820
             ||  :||||| |  |||:|:  |  ||:|  ::        ||||:  |||||:|||
Db   745 GHTDIVTLLLKNGASPNEVSSDGTTPLAIA------KRLGYISVTDVLKVVTDETSFVLVS 799

Qy   821 GDHNILFDQ 829
             .| : |
Db   800 DKHRMSFPE 808

RESULT 9
ADD47763
ID   ADD47763 standard; protein; 1881 AA.
XX
AC   ADD47763;
XX
DT   02-DEC-2004  (revised)
DT   29-JAN-2004  (first entry)
XX
DE   Human Protein XP_016774, SEQ ID NO 13459.
XX
KW   Human; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
OS   Unidentified.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C, D'urso D, Befort K, Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; XP_016774.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Example 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
```

CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human protein (described in Table 3
CC   of the specification) which is differentially expressed during pain.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1881 AA;

Query Match   9.1%;   Score 390.5; DB 7;   Length 1881;
Best Local Similarity   26.0%;   Pred. No. 1.3e-20;
Matches   174;   Conservative 100;   Mismatches 246;   Indels 149;   Gaps   31;

Qy   194 RIVASKGDLSEDQVEBILDIIFMNESEIAEGISNPLHADVDNNPVK--------GA----- 241
                 ::||| :::| ::|
Db   256 RULLDRGAQIETKTKDEL-------------------------TPLHCAARNGHVRISEILLDHGAPIQA 300

Qy   242 --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT 298
             :| |: |||
Db   301 KTKNGLSPIHMAAQGDHLD--CVRLLLQYDAEIDDI------TLD--HLTPLHVAAHCG 349

Qy   299 GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK 357
             ::| ::|:|::: || :
Db   350 HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL 405

Qy   358 NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT-IAHYAVKNK 407
             |||| ||:| |:| :|:||::|||
Db   406 TPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK 460

Qy   408 NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP 467
             :::   ::| || |:::|:
Db   461 ------------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP 506

Qy   468 LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK 525
             ||| :|:||:| |: :|| |||||| || |
Db   507 LHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA 565

Qy   526 GEDTVAKSLLMEDYKGFTPLHPVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE 585
             |:: ||:
Db   566 GKN-----------GLTPLHVAVHHNNLDIVKL-------------------------- 587

Qy   586 LVNLGDFPESENILGSPNAKFLQHIQSANFGFSPAARRGIVSSNHNVMKDILNFVGDSLHLP 645
             |:| |:|: : |::|| :|| :|||:: :
Db   588 LLPRG------GSP------HSPAWN-GYTPLHIAAKQNOVEVARSLLQY-GGSANAE 631

Qy   646 SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPPLNLACLRGDNEVVRGLVG 705
             :| ||| |: |:|:: : ||| :| |::|:|||
Db   632 SVQGVTPLHLAAQEGHAEMVALL--SKQANGNLGNKSGLTPLHLVAQEGHVPVADVLI- 688

Qy   706 QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG 763
             |:| ||| ::|: :||||| |:: |:| ||||: :|
Db   689 KHGVMVDATTRMG---YTPLHVASHYGYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ 744

Qy   764 GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL---LDRGADINLD 820
             |:|: ||| :|| :::|:|:| ::|||:||:||:|:||| |:||::|||
Db   745 GHTDIVTLLKNGASPNEVSSDGTTPLAIA-----KRLGYISVTDVLKVVTDETSFVLVS 799

Qy   821 GDNILFDQ 829
             : | :|
Db   800 DKHRMSFPE 808

RESULT 10
ABM80426
ID   ABM80426 standard; protein; 899 AA.
XX
AC   ABM80426;
XX
DT   18-NOV-2004 (first entry)
XX
DE   Tumour-associated antigenic target (TAT) polypeptide PRO80961, SEQ:1071.
XX
KW   Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW   tumour; diagnosis; cell proliferative disorder; breast cancer;
KW   colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW   central nervous system cancer; bladder cancer; pancreatic cancer;
KW   cervical cancer; melanoma; leukaemia; hybridisation probe;
KW   chromosome identification; chromosome mapping; gene mapping;
KW   gene therapy; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200403615-A2.
XX
PD   15-APR-2004.
XX
PF   29-SEP-2003; 2003WO-US028547.
XX
PR   02-OCT-2002; 2002US-0414971P.
XX
PA   (GETH) GENENTECH INC.
XX
PI   Wu TD, Zhang Z, Zhou Y;
XX
DR   WPI; 2004-347921/32.
DR   N-PSDB; ACN37888.
XX
PT   New tumor-associated antigenic target polypeptides and nucleic acids,
PT   useful in preparing a medicament for treating or detecting a
PT   proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT   prostate cancer or tumor.
XX
PS   Claim 12; SEQ ID NO 1071; 7272pp; English.
XX
CC   The invention relates to human tumour-associated antigenic target (TAT)
CC   polypeptides, and their related nucleic acids. The TAT polypeptides are
CC   overexpressed in cancer tissues compared to normal tissues, and may thus
CC   serve as effective targets for the diagnosis and treatment of cancer in
CC   mammals. The invention also relates to nucleic acid and polypeptide
CC   sequences at least 80% identical to the TAT nucleic acids and
CC   polypeptides; expression vectors and host cells comprising a TAT nucleic
CC   acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC   molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC   TAT polypeptide; and methods and compositions for the treatment or
CC   diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC   antibodies, antagonists, binding molecules and compositions are useful
CC   for diagnosing or treating a cell proliferative disorder associated with
CC   increased TAT expression, particularly cancers such as breast cancer,
CC   colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC   cancer, pancreatic cancer, cervical cancer, cancers of the central
CC   nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC   used as hybridisation probes, in chromosome and gene mapping, in
CC   chromosome identification and in gene therapy. The present sequence
CC   represents a TAT polypeptide of the invention
XX
SQ   Sequence 899 AA;

```
Query Match         8.9%;    Score 382.5;  DB 8;  Length 899;
Best Local Similarity  25.8%;    Pred. No. 1.7e-20;
Matches 163;   Conservative  91;  Mismatches 246;   Indels 131;  Gaps  24;

Qy   209  EILDIIFMNESEIA---EGISNPLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALG  265
                :::.:.::.    :  ::. ::        :.:  ::: ::.: ::.  .:.
Db    33  EVKLLVSHGAEVTCKDKKSYTPLHAAASSGMISVVKYLL-------------------   72

Qy   266  EVENDEGDLGA-----NAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANPA  320
                ::::::        ::          ::::::  ::
Db    73  -------DLGVDMNEPNAY------GNTPLHVACYN--------GQDVVVNELIDCGAIV  111

Qy   321  NQ------SPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKG  374
                         ::::   :.    .  ::: :::::        :::
Db   112  NQKNEKGFTPLHFAAASTH-GALCLELLVGNGADVNMKSKDGKTPLHMTALH-------G  163

Qy   375  VMDHVKNSSEFQDLVNDPPDYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAP  434
                ::. :.   ::      :.  :::.::::.:::::::::::::::::.:::::::
Db   164  RFSRSQTIIQSGAVIDCEDKGNGTPLHIAARYGHE--LLINTLITSGADTAKRGIHGMFP  221

Qy   435  IHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQ  494
                ::.:.::::  .  :::::::::::::::::::::::::::::.:::::::::
Db   222  LHLAALSGFSDCCRKLLSSGFDIDTPDDFGRTCLHAAAAGGNLECLNLLLN-TGADFNKK  280

Qy   495  NNDGVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGGGSR  554
                  :::::::::::::.:  : : :::   :::.     ::::::::::::
Db   281  DKFGRSPLHYAAANCNYQCLFALVGSGASVNDLDE-------RGCTPLHYAA---TS    327

Qy   555  DTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANF  614
                ::                   : ::.    :.::         .   :
Db   328  DT-------------DGKCLEYLLRNDANPG----IRDKQGYNAV-------H-YSAAY  361

Qy   615  GFSPARRGIVSSNNNVMXDIL------NFVGDSLHLPSERGYNAMQVAALFGDKEAVKML  668
                                    :: : .:.    ::::  .:.
Db   362  GHRLCLQLIAS--ETPLDVLMETSGTDMLSDS---DNRATISPHLRAYHGHQALEVL   415

Qy   669  AKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGIDINQRMGGSDKNTVLHYAIS  728
                :::     :.: .::::: .:::: :.:::::::::
Db   416  VQSL--LDLDVRNSSGRTPLDLAAPKGHVECVDVLINQ-GASILVKDYILKRTPIHAAAT  472

Qy   729  KGDSFIVQKILAH----TGVDVNCENNLGQTPLHIAVEGGDPKIVSSLLKAGAVVNRLDD  784
                :           :::::::  ::::::::::          :: :   :
Db   473  NGHSECLRLLIGNAEPQNAVDIQDGN--GQTPLMLSVLNGHTDCVYSLLNKGANVDAKDK  530

Qy   785  NGRSVLSSAIVPGRKEKGVLGIVNKLLDRGA  815
                ::   :.:::           :::: :
Db   531  WGRTALHRGAVTGHEE-----CVDALLQHGA  556
```

```
RESULT 11
ABU11523
ID   ABU11523 standard; protein; 1536 AA.
XX
AC   ABU11523;
XX
DT   12-FEB-2003  (first entry)
XX
DE   Human MDDT polypeptide SEQ ID 470.
XX
KW   MDDT; human; disease detection and treatment molecule polypeptide;
KW   anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
KW   haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;
KW   gene therapy; protein replacement therapy; cell proliferative disorder;
KW   cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;
KW   anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
KW   Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
KW   psoriasis; hepatitis.
XX
OS   Homo sapiens.
XX
PN   WO200279449-A2.
XX
```

```
PD   10-OCT-2002.
XX
PF   27-MAR-2002;  2002WO-US009944.
XX
PR   28-MAR-2001;  200lUS-0279619P.
PR   29-MAR-2001;  200lUS-0280067P.
PR   29-MAR-2001;  200lUS-0280068P.
PR   16-MAY-2001;  200lUS-0291280P.
PR   17-MAY-2001;  200lUS-0291829P.
PR   17-MAY-2001;  200lUS-0291849P.
PR   19-JUN-2001;  200lUS-0299428P.
PR   20-JUN-2001;  200lUS-0299776P.
PR   20-JUN-2001;  200lUS-0300001P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI   Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;
PI   Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI   Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI   Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX
DR   WPI; 2003-058431/05.
DR   N-PSDB; ABX34513.
XX
PT   New purified disease detection and treatment molecule proteins and
PT   polynucleotides, useful for diagnosing, treating or preventing cancers
PT   (e.g. leukaemia or sarcoma), anaemia, Crohn's disease, AIDS, osteoporosis
PT   or hepatitis.
XX
PS   Claim 27; SEQ ID NO 470; 339pp + Sequence Listing; English.
XX
CC   This invention describes a novel disease detection and treatment molecule
CC   polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,
CC   osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,
CC   antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides
CC   and the polypeptide of the invention can be used for gene therapy,
CC   protein replacement therapy and are useful for treating a variety of
CC   diseases or conditions. These polypeptides or polynucleotides are
CC   particularly useful for diagnosing, treating or preventing cell
CC   proliferative disorders (e.g. cancers including adenocarcinoma,
CC   leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's
CC   disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
CC   syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or
CC   hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded
CC   by ABU11450-ABU11845, described in the disclosure of the invention. NOTE:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1536 AA;
```

```
Query Match         8.9%;    Score 382.5;  DB 6;  Length 1536;
Best Local Similarity  27.7%;    Pred. No. 4e-20;
Matches 156;   Conservative  84;  Mismatches 206;   Indels 117;  Gaps  26;

Qy   289  PLHHAAKNCTGDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNL  347
                :::::   :::   : :.:: :.     :        :         :::::
Db    68  PLHVAA-HCGHHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASI  124

Qy   348  DFGLVDGDGKNPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT  398
                ::  : :::::::      ::::    :::           :           :::
Db   125  D--AVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHT  177

Qy   399  -IAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDV  457
                :: ::::  :       :      ::  :        :::::
Db   178  EVAKYLLQNK-----------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANP  223

Qy   458  NSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL  517
                  :::::::::: : :  :::::::::  :::::     :  ::::::::::.: ::
Db   224  NLATTAGHTPLHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELL  282

Qy   518  --GSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRA  575
```

```
Db  283 LERDAHPNAAGKN---------GLTPLHVAVHHNNLDIVKL----------- 314
            | |  ||  |::              |||||
Qy  576 ALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSP---ARRGIVSSNHNVM 631
            |:  |        :||:||  |           |:: |:|
Db  315 ---------LLPRG--------GSP------HSPAWN-GYTPLHIAAKQNQVEVAHSLL 349
Qy  632 KDILNFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLA 691
            ||||| :|||||||||||||::|:||
Db  350 Q-----YGGSANAESVQGVTPLHLAAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHLV 402
Qy  692 CLRGDNEVVRGLVGQHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCE 749
            :|: |           |||||          |||:|
Db  403 AQEGHVPVADVLI-KHGVMVDATTRMG--YTPLHVASHYGNIKLVKFLLQHQ-ADVNAK 457
Qy  750 NNLGQTPLHLAVEGGDPKIVSSLLKAGAVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNK 809
            ||:||| |||| |||  |||: ||||||||||||||           :   |: :
Db  458 TKLGYSPLHQAAQQGHTDIVTLLLKNGASPNEVSSDGTTPLAIA-----KRLGYISVTDV 512
Qy  810 L--LDRGADINLDGDHNILFPQ 829
            |   |||        |:|  |:
Db  513 LKVVTDETSFVLVSDKHRMSFPE 535
```

RESULT 12
ABG00972
ID ABG00972 standard; protein; 4274 AA.
XX
AC ABG00972;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #963.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO20175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PT WPI; 2001-639362/73.
DR N-PSDB: AAS65159.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 31331; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4274 AA;

Query Match   8.8%; Score 377.5; DB 4; Length 4274;
Best Local Similarity 25.5%; Pred. No. 4.8e-19;
Matches 175; Conservative 92; Mismatches 277; Indels 141; Gaps 26;

```
Qy  184 TSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNES----------EIAEGISNPLH 230
            | ||  ||  ||                                 :|  |||| |
Db  197 TKGKVRLPLAHIIAARKDD------------TKAAALLQNDNNADVESKSGFT-PLHIAAHYGNIN 249
Qy  231 ------------ADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNV 281
               :|:  :|:|  ||  :|::
Db  250 VATLLLNRAAAVDFT---ARNDITPLH-VASKRGNANMVKLL---------------LDRGAK-IDA 296
Qy  282 LDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEE 341
            :|    |||  |  ||||   ||                         |  |
Db  297 KTRDGLTPLHCGARSGHEQVVEMLLDRAAPILSKTKN--GLSPLHMATQGDHLNCVQL-- 352
Qy  342 FTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAH 401
            :    |  |   |  |||| ||           :
Db  353 LLQHNVPVDDVTNDYLTALHVAAHCGHYKVAKVLLDKKANP-------NAKALNGFTPLH 405
Qy  402 YAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQD 461
            :|  ||  :|  ::|  |||    :||:   ||||||:||||
Db  406 IACK-KNR-IKVMELLLKHGASIQAVTRSGLTPIHVAAPMGHVNIVSQLMHHGASPNTTN 463
Qy  462 VNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL-GS 519
            | | ||||:|  |:||  |   |:|| || : |::| ||| :||| |  ::|  |
Db  464 VRGETALHMAARSGQAE-VVRYIVQDGAQVEAKAKDQTPLHISARLGKADIVQQLLQQG 522
Qy  520 AKPNI------------KGEDTVA-------KSLLMEDYKGFTPLHFVAGGGSRDTFR 558
            |:|:             ||| |        |:  ::|:||||||| ||
Db  523 ASPNAATTSGYTPLHLSAREGHEDVAAFLLDRGASLSITTKKGFTPLHVAA-------- 573
Qy  559 VVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENIL---GSPNAKFLQHIQSANF 614
            ||||:           :|: |  |||| ||||  ||                :
Db  574 --------------------KYGKLEVANLLLQKSASPDA------------AGKS 597
Qy  615 GFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAKSAKP 674
            | :|      |::|  :|||||||||||||||||||||||||:|||||||:|||||:|
Db  598 GLTPLHVAAHYDNQKVALLLLD-QGASPHAAAKNGYTPLHIAAKKNQMDIATTLLEYG-- 654
Qy  675 SDLNFKTSATPTPLNLACLRGDNEVVRGLVGQH-GIDINQRMGSDKNTVLHYAISKGDSF 733
            :|  :|:  |                :||||   |||         |:|:
Db  655 ADANAVTRQGIASVHLAAQEGHVDMVSLLLGRNANVALSNKSGL--TPLHLA-AQEDRV 710
Qy  734 LVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVNRLDDNGRSVLSSA 793
            :: |:||| ||||: :  ||:||||:|| |  :||||:||| ||||||||||||||:|
Db  711 NVAEVLVNQGAHVDAQTKMGYTPLHVGCHYGNIKIVNFLLQHSAKVNAKTKNGYTPLHQA 770
Qy  794 IVPGRKEKGVLGIVNKLLDRGADIN 818
            :: |||:||:||:|:|||||||:|
Db  771 -----AQQGHTHINVLLQNNASPN 790
```

RESULT 13
ABG07375
ID ABG07375 standard; protein; 4386 AA.
XX
AC ABG07375;
XX
DT 13-FEB-2002 (first entry)

```
Db   360  LLQHNVPVDDVTNDYLTALHVAAHCGHYKVAKVLLDKKANP------NAKALNGFTPLH  412

Qy   402  YAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQD  461
               |:|:|  |||     : |   :|::      |:  ||     |:    |:|:
Db   413  IACK-KNR-IKVMELLLKHGASIQAVTESGLTPIHVAAFMGHVNIVSQLMHHGASPNTTN  470

Qy   462  VNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GS  519
               |:|:||  |:|  | ::  |:|  ||   |  :| | ||:  ||: :::|
Db   471  VRGETALHMAARSQQAB-VVRYLVQDGAQVEAKAKDDQTPLHISARLGKADIVQQLLQQG  529

Qy   520  AKPNI----------KGEDTVA-----KSLLMEDYKGFTPLHFVAGGSRDTFR  558
               ||:|          ||  |:        :::  |:  ||    ||:
Db   530  ASPNAATTSGYTPLHLSAREGHEDVAAFLLDHGASLSITTKKGFTPLHVAA---------  580

Qy   559  VVRKNYEKCHDLATTIRAALMQDRSGGELVNLGDFESENLL-----GSPNAKFLQHIQSANF  614
               ||  :|:  |  |  |:|    :|:|:            ||:||||||:||||:
Db   581  -------------------------KYGKLEVANLLLQKSASPDA------AGKS  604

Qy   615  GFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAKSAKP  674
               |:   ||::|:||||| ::|:   ||:||||: ::|: |:::|:|||:|:|:|
Db   605  GLTPLHVAAHYDNQKVALLLLD-QGASPHAAAKNGYTPLHIAAKKNQMDIATTLLEYG--  661

Qy   675  SDLNFKTSATPTPLNLACLRGDNEVVRGLVGQH-GIDINQRMGSDKNTVLHYAISKGDSF  733
               :|:|    |||:|: |:|:  ||::|::|   ::||  |    |::|||:||::|:
Db   662  ADANAVTRQGIASVHLAAQEGHVDMVSLLGRNANVNLSNKSGL---TPLHLA-AQEDRV  717

Qy   734  LVQKILAHTGVDVVNCENNLGQTPLHLAVEGGHPDKIVSSLLKAGAVVNRLDDNGRSVLSSA  793
               ||:|  | :|||:||  |||:|||:|||:  ::||: :||| || ||: ||:||||:|:|
Db   718  NVAEVLVNQGAHVDAQTKMGYTPLHVGCYGNIKIVNFLLQHSAKVNAKTKNGYTPLHQA  777

Qy   794  IVPGRKEKGVLGIVNKLLDRGADIN  818
               :|    |    ||:|||||:
Db   778  -----AQQGHTHINVLLQNNASPN  797
```

RESULT 14
ADR08749

```
Query Match             8.8%;  Score 377.5; DB 4; Length 4386;
Best Local Similarity   25.5%;  Pred. No. 5e-19;
Matches  175; Conservative  92; Mismatches  277; Indels  141; Gaps   26;

Qy   184  TSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNES----EIAEGISNPLH-------  230
               |  ||:   |:   ::| ||:::  |  :|   |:   :|:|:|:||
Db   204  TKGKVRLPALHIAARKDD-----TKAAALLQNDNNADVESKSGFT-PLHIAAHYGNIN  256

Qy   231  --------ADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNV  281
                       ||  ||:| :||:|  |:|:   ||  |:    :  | | |:
Db   257  VATLLLNRAAAVDFT---ARNDITPLH-VASKRGNANMVKLL------LDRGAK-IDA  303

Qy   282  LDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEE  341
               |: | ||||:  : | |  ||: | |::  | |||||:||:|       |:::|
Db   304  KTRDGLTPLHFCGARSGHEQVVEMLLDRAAPILSKTKN--GLSPLHMATQGDHLNCVQL--  359

Qy   342  FTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFODLVNDPDYFGNTIAH  401
               |  | |    |||   :|||:||| |  ||| |::|  ::   |:   || ::| | |
```

```
XX
PS   Claim 1; SEQ ID NO 2255; 2686pp; English.
XX
CC   This invention relates to novel, isolated full length human cDNA
CC   molecules and the encoded proteins thereof. Specifically, it refers to
CC   cDNA cloned obtained by an oligo-capping method, where none of these
CC   clones are identical to any known human mRNAs. The present invention
CC   describes in immunoassay to identify agonists and antagonists, as well as
CC   antibodies, antisense molecules and siRNAs that can all be used to bind
CC   to and modulate expression of the cDNA molecules. As such, these
CC   molecules are useful for diagnostic markers or therapeutic targets for
CC   the various diseases or morbid states. In particular, they are useful in
CC   gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC   disease, Parkinson's disease, dementia, short memory and various cancers,
CC   as well as for maintaining equilibrium of sense or motor function, and
CC   for treating emotional reaction, fear response and panic. Accordingly,
CC   they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC   cytostatic and tranquiliser activities. This polypeptide is a protein
CC   encoded by a full length human cDNA sequence of the invention. NOTE: This
CC   sequence is not given in the sequence listing of the specification but
CC   can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
CC   office.
XX
SQ   Sequence 687 AA;

Query Match             8.7%;  Score 373.5;  DB 8;  Length 687;
Best Local Similarity  24.0%;  Pred. No. 5.9e-20;
Matches  190;  Conservative 115;  Mismatches 299;  Indels 187;  Gaps 35;

Qy     72 VEDDEDLPPTLPPRTFSGEGYDDVGVSMPTVSRGIYOPP----IVQDSNLYSSI-GGVPQ 126
             :|::||                  :||   ::  |  |    :|: |   |
Db      9 LEEVEDESPAFISK------------------LPQENKSLHSPPSGNVLVRYPSLVQAIFNGDPD 55

Qy    127 EAQYDAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDVRNAVLGNKEIKE------ 179
              |:                         |||||          :|||:|  |
Db     56 EV------------RALIFK----KEDVNFQDNEKRTPLHAA--AYLGDAEIIELLILSGA 98

Qy    180 ------EWLTTSGPVRDIADRIVASKGDLSEDQVEBILDIIFMNESSEIAEGISN---PL 229
               ||      |:|             :    ::        :     |
Db     99 RVNAKDSKWLT--PLH---RAVASCS------EEAVQVLLKHSADVNARDKNWQTPL 144

Qy    230 HADVDNNPVKGAKNVMTLMHLYYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLP 289
            ||| ||:|                   |  :|           |||
Db    145 HIAAANKAVKCAEALVPLLSNV------------------------NVSDRAGRTA 176

Qy    290 LHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEFTSRNLDF 349
            ||||: ||  |   :| :| ::  :  :  :       |     :
Db    177 LHHAAFSGHGEMVKLLLSRGAN--INAFDKKDRRAIHWAAYMGHIEVVKL--LVSHGAEV 232

Qy    350 GLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPPDYFGNTIAHYAVKNKNA 409
            |:||||||        ||||:| :|                   :| :   |:| |||
Db    233 TCKDKKSYTPLHAAAS------------SGMISVVKYLLDLGVDMNEPNAYGNTPLHVACYN-GQ 284

Qy    410 DLTLFNMLKASGADLNVRNVVGRAPIH-VASSNGKANAVSGLVSCGIDVNSQDVNGDTPL 468
            | :|  |   | | |    |||    || :|  | |||         |:||     ||
Db    285 DVVV-NELIDCGAIVNQKNEKGFTPLHFAAAASTHGALCLELLVGNGADVNMKSKDGKTPL 343

Qy    469 HIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKALGSAKPNIKGED 528
            |:|        ::  | ::| :  |||||  |  |||:||  :        |
Db    344 HMTALHGRFSRSQTII-QSGAVIDCEDKNGNTPLHIAARYGHELLINTLITS------GAD 397

Qy    529 TVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATI----RAALMQDRSGG 584
            |                :|  || |           |      |     |||
Db    398 TAKRGI------HGMPLHLAALSGFSD---CCRKLLSSGFDIDTPDDFGRTCLHAAAAGG 449

Qy    585 ELVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILN---FVGD- 640
            |                |   |:        :       |        :|    ||||
Db    450 NL------ECLNLLLNTGADF--NKKDKFGRSPLHYAAANCNYQCLFALVGSGASVNDL 500

Qy    641 --SLH------------LPSERGYNAMQVAALFGDKEAVKMLAKS 671
              :||              :|||| |  :|||:|| ::||||::
Db    501 DERGCTPLHYAATSDTDGKCLEYLLRNDANPGVRDKQGYNAVHYSAAYGHRLCLQLIA-S 559

Qy    672 AKPSDLNFKTS-----------------ATPTPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDKN 720
             |::|||||||: ||                :|: |:|:||         || || :||
Db    560 ETPLDVLMETSGTDMLSDSDNRATISPLHLAAYHGHHQALEVLV-QSLLDLDVRNSSGR- 617

Qy    721 TVLHYAISKGDSPLVQKILAHTGVDVNCENN-LGQTPLHLAVEGGDPKIVSSLLKAGAVV 779
            |::| |:||  ||:|::| |: ||||   :  ::||||||||| :: ::|| ||:|| ||
Db    618 TPLDLAAFKGHVECVD-VLINQGASILVKDYILKRTPIHAAATNGHSECLRLLIGNAEPQ 676

Qy    780 NRL---DDNGR 787
            |:|   : ||
Db    677 NAVDIQDGNGQ 687


RESULT 15
AAE33684
ID   AAE33684 standard; protein; 1056 AA.
XX
AC   AAE33684;
XX
DT   16-APR-2003  (first entry)
XX
DE   Human structural and cytoskeleton-associated protein (SCAP) #18.
XX
KW   Human; structural and cytoskeleton-associated protein; SCAP; leukaemia;
KW   cell proliferative disorder; actinic keratosis; cancer; gastroenteritis;
KW   cirrhosis; psoriasis; adenocarcinoma; Creutzfeldt-Jakob disease; stroke;
KW   sarcoma; neurological disorder; epilepsy; seasonal affective disorder;
KW   Huntington's disease; Alzheimer's disease; lymphoma; melanoma; myeloma;
KW   anxiety; schizophrenia; amnesia; viral infection; pneumonia; influenza;
KW   arteriosclerosis; smallpox; Colorado tick fever; rabies; gene therapy;
KW   neuroprotective; nootropic; neuroleptic; cytostatic; virucide.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Binding-site    867..874
FT                   /note="ATP/GP-binding site motif A"
XX
PN   WO200210009-A2.
XX
PD   19-DEC-2002.
XX
PF   06-JUN-2002;  2002WO-US017956.
XX
PR   07-JUN-2001;  2001US-0296865P.
PR   08-JUN-2001;  2001US-0296878P.
PR   15-JUN-2001;  2001US-0298664P.
PR   21-JUN-2001;  2001US-0300149P.
PR   29-JUN-2001;  2001US-0302340P.
PR   06-JUL-2001;  2001US-0303481P.
PR   12-JUL-2001;  2001US-0305059P.
PR   21-DEC-2001;  2001US-0343557P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Tang TY,    Warren BA,  Honchell CD,  Richardson TW,  Elliott VS;
PI   Walia NK,   Yue H,  Batra S,  Griffin JA,  Baughn MR,  Forsythe IJ;
PI   Burford N,  Emerling BM,  Sanjanwala MM,  Khan FA,  Lu DAM,  Hafalia AJA;
PI   Nguyen DB,  Yang J,  Li JX,  Becha SD,  Yao MG,  Gietzen KJ,  Luo W;
PI   Lee EA,     Ison CH,  Lasek AKW;
XX
DR   WPI; 2003-148791/14.
DR   N-PSDB; AAD51581.
XX
PT   New human structural and cytoskeleton-associated proteins and genes,
PT   useful for diagnosing or treating cancers (e.g. leukemia or lymphoma),
PT   viral infections (e.g. influenza) or neurological disorders (e.g.
PT   epilepsy or stroke).
XX
PS   Claim 1; Col 195-198; 121pp; English.
XX
CC   The invention relates to human structural and cytoskeleton-associated
CC   proteins (SCAP) and genes. SCAP sequence and agonist are useful for
```

```
CC      treating a disease or condition associated with decreased expression of
CC      functional SCAP. The antagonist is useful for treating a disease or
CC      condition associated with over expression of functional SCAP. The
CC      antibody that specifically binds to the polypeptide is useful for
CC      diagnosing a condition or disease associated with the expression of SCAP.
CC      SCAP sequence, agonists and antagonists are particularly useful for
CC      diagnosing, treating or preventing cell proliferative disorders (e.g.
CC      actinic keratosis, arteriosclerosis, cirrhosis, psoriasis, or cancers
CC      including adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma or
CC      sarcoma), neurological disorders (e.g. epilepsy, seasonal affective
CC      disorder, Huntington's disease, stroke, Alzheimer's disease, anxiety,
CC      Creutzfeldt-Jakob disease, schizophrenia or amnesia) or viral infections
CC      (e.g. pneumonia, herpes, influenza, Colorado tick fever, smallpox, rabies
CC      or gastroenteritis). SCAP DNA is also used in gene therapy. The present
CC      sequence is human SCAP protein
XX
SQ      Sequence 1056 AA;

Query Match             8.7%;      Score 373.5;  DB 6;    Length 1056;
Best Local Similarity  23.9%;      Pred. No. 1.1e-19;
Matches   198; Conservative  101; Mismatches  303;   Indels  227;   Gaps    35;

Qy        108 QPPIVQDSNLYSSIGGVPQEAQYDAAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDV 167
              ||||:||              : ::        :||||
Db          9 QPPLVQ-----AIFSGDPEI----------RMLIHK---TEDVNTLDSEKRTPLHVA 48

Qy        168 RNAVLGNKEIKE-----------------EWLTTSGPVRDIADRIVASKGDLSEDQVEEILDI 213
                  ||||              |:::  :  :||:|  ||| :|||||:|||||
Db         49 --AFLGDAEIIELLILSGARVNAKDNMWLT---PLH----RAVASRS------EEAVQV 92

Qy        214 IFMNESEIAEGISN--PLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIV---KAL-GE 266
              : ||:|             :|  : ::::|  :|  |||  |||| |||||
Db         93 LIKHSADVNARDKNWQTPLHVAAANKAVKCCAEVIIPLLSSVNVSDRGGRTALHHAALNGH 152

Qy        267 VE--------------NDEGDLGANAY------------------NVLDSEGNLPLHH 292
                                :                         ||||:|:|||||
Db        153 VEMVNLLLAKGANINAFDKKDRRALHWAAYMGHLDVVALLINHGAEVTCKDKKGYTPLHA 212

Qy        293 AAKNCTGDKLK------------------LC---MEKTKTDFIDTANFANQ------ 322
              ||||||||:|                   ||   ||: : ||:|||||||
Db        213 AASNGQINVVKGHLLNLNLGVEIDEINVYGNTALHIACYNGQDAVVNELIDYGANVNQPNNNG 272

Qy        323 -SPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDH--- 378
               :||||||            |||:||||:  ||||  ||||||| |  :  :  |:
Db        273 FTPLHFAAASTH-GALCLELLVNNGADVNIQSKDGKSPLHMTAVH-----------GRFTRSQT 324

Qy        379 -VKNSSEFQDLVNDPDYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHV 437
               :||              : :    |||||:| |||||:|||| |||| ||||||:|
Db        325 LIQNGGEIDCVDKD----GNTPLHVAARYGHE--LLINTLITSGADTAKCGIHSMFPLHL 378

Qy        438 ASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNND 497
              ||:||            : ||||||||||||||||  ::::::::::|| :
Db        379 AALNAHSDCCRKLLSSGFEIDTPDKFPGRTCLHAAAAGGNVECI-KLLQSSGADFHKKDKC 437

Qy        498 GVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAG------- 550
              ||:||||::|          ::|:              :: :|||||
Db        438 GRTPLHYAAANCHFHCIETLVTTGANVNETD--------DW-GRTALHYAAASDMDRNK 487

Qy        551 ---GGSRDTFRVVRKNYEKCHDLATIRAAL----------MQDRSGGELVN----LG 590
                    :::|  : ::: :||                     : ::: ::  :
Db        488 TILGNAHDN---SEELERARELKEKEATLCLEFLLQNDANPSIRDKEGYNSIHYAAAYG 543

Qy        591 DFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDIILNFVGDSLIHLPSERGY 650
               :||::||     :|  |:||| ::: : :|           :|| |:   ||||
Db        544 HRQCLELL----------LERTNSGFPEESDSGATKS--------------PLHL- 573

Qy        651 NAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGID 710
              : : |::         :|||: ::|||||  : :||| | ||||| ||:|||| |||:|
Db        574 -----AAYNGHHQALEVLLQSL--VDLDIRDEKGRTALDLAAFKGHTECVEALINQ-GAS 625

Qy        711 INQRMGSDKNTVLHYAISKGDSFLVQKIL--AHTGVDVNCENNLGQTPLHLAVEGGDPKI 768
              : : | :||||:|||| :::| ||:::    :|||| || ||: |||||||||:| |||
Db        626 IFVKDNVTKRTPLHASVINGHTLCLRLLLEIADNPEAVDVKDAKGQTPLMLAVAYGHIDA 685
```

```
Qy        769 VSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADI 817
              ||:||||:||||:|::|:||| |||||    :|||||||||||||||||
Db        686 VSLLLEKEANVDTVDILGCTALHRGIMTGHEE----CVQMLLEQEVSI 729
```

```
Search completed: February 26, 2005, 18:50:50
Job time : 180 secs
```

```
                    GenCore version 5.1.6
         Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      February 26, 2005, 18:47:23 ; Search time 44 Seconds
                                 (without alignments)
                                 1409.849 Million cell updates/sec

Title:           US-10-701-038-2
Perfect score:   4310
Sequence:        1 PSGDIQDSQQDQQEQDQQQ.........DRGADINLDGDHNILFDQCL 831

Scoring table:   BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:        513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters:          513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued Patents AA:*
                 1: /cgn2_6/ptcdata/1/iaa/5A_COMB.pep:*
                 2: /cgn2_6/ptcdata/1/iaa/5B_COMB.pep:*
                 3: /cgn2_6/ptcdata/1/iaa/6A_COMB.pep:*
                 4: /cgn2_6/ptcdata/1/iaa/6B_COMB.pep:*
                 5: /cgn2_6/ptcdata/1/iaa/PCTUS_COMB.pep:*
                 6: /cgn2_6/ptcdata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	Query Match %	Match Length	DB	ID	Description
1	1963	45.5	390	3	US-09-561-322-2	Sequence 2, Appli
2	390.5	9.1	1719	4	US-09-949-016-6966	Sequence 6966, Ap
3	390.5	9.1	1856	4	US-09-949-016-6964	Sequence 6964, Ap
4	390.5	9.1	1880	4	US-09-949-016-5876	Sequence 5876, Ap
5	390.5	9.1	1881	4	US-09-949-016-6965	Sequence 6965, Ap
6	390.5	9.1	1883	4	US-09-949-016-9010	Sequence 9010, Ap
7	390.5	9.1	1883	4	US-09-949-016-9011	Sequence 9011, Ap
8	390.5	9.1	1883	4	US-09-949-016-9013	Sequence 9013, Ap
9	390.5	9.1	1883	4	US-09-949-016-9012	Sequence 9012, Ap
10	390.5	9.1	1883	4	US-09-949-016-9014	Sequence 9014, Ap
11	390.5	9.1	1883	4	US-09-949-016-9015	Sequence 9015, Ap
12	390.5	9.1	1883	4	US-09-949-016-9016	Sequence 9016, Ap
13	390.5	9.1	1883	4	US-09-949-016-9017	Sequence 9017, Ap
14	377.5	8.8	4377	4	US-09-949-016-6978	Sequence 6978, Ap
15	370.5	8.6	994	4	US-10-164-595-38	Sequence 38, Appli
16	368.5	8.5	1839	2	US-09-172-977-4	Sequence 4, Appli
17	368.5	8.5	1839	4	US-10-404-108-4	Sequence 4, Appli
18	368.5	8.5	2753	4	US-09-949-016-7659	Sequence 7659, Ap
19	368.5	8.5	2753	4	US-09-949-016-7660	Sequence 7660, Ap
20	368.5	8.5	3924	4	US-09-538-092-1246	Sequence 1246, Ap
21	368	8.5	1619	4	US-09-392-812A-4	Sequence 4, Appli
22	355.5	8.2	843	2	US-09-172-977-3	Sequence 3, Appli
23	355.5	8.2	843	4	US-10-404-108-3	Sequence 3, Appli
24	351	8.1	1709	4	US-09-392-812A-6	Sequence 6, Appli
25	350	8.1	1088	3	US-09-082-059-2	Sequence 2, Appli
26	332	7.7	683	4	US-10-164-595-71	Sequence 71, Appl
27	324.5	7.5	3913	4	US-09-949-016-10933	Sequence 10933, A
28	324	7.5			US-09-392-812A-2	Sequence 2, Appli
29	321	7.4		2	US-09-031-485-33	Sequence 33, Appl
30	321	7.4		4	US-08-847-429A-33	Sequence 33, Appl
31	321	7.4		3	US-09-065-474-33	Sequence 33, Appl
32	321	7.4		3	US-09-557-034-33	Sequence 33, Appl
33	320	7.4		4	US-09-196-387-10	Sequence 10, Appl
34	320	7.4		4	US-09-841-835-10	Sequence 10, Appl
35	320	7.4		3	US-09-196-387-2	Sequence 2, Appli
36	320	7.4		3	US-09-841-835-2	Sequence 2, Appli
37	320	7.4		4	US-09-972-115A-8	Sequence 8, Appli
38	300	7.0		4	US-09-509-802-2	Sequence 2, Appli
39	300	7.0		4	US-09-188-930-334	Sequence 334, App
40	300	7.0		4	US-09-312-283C-334	Sequence 334, App
41	289.5	6.7		3	US-09-439-313-378	Sequence 378, App
42	289.5	6.7		3	US-09-352-616A-378	Sequence 378, App
43	289.5	6.7		4	US-09-636-215-378	Sequence 378, App
44	289.5	6.7		4	US-09-685-166A-378	Sequence 378, App
45	289.5	6.7		4	US-09-679-426-378	Sequence 378, App

```
                              ALIGNMENTS

RESULT 1
US-09-561-322-2
; Sequence 2, Application US/09561322
; Patent No. 6355777
; GENERAL INFORMATION:
;  APPLICANT:  Walker, David H.
;  APPLICANT:  McBride, Jere W.
;  TITLE OF INVENTION:  P43 Antigen for the Immunodiagnosis of Canine Ehrlichiosis and
;  FILE REFERENCE:  D6325
;  CURRENT APPLICATION NUMBER: US/09/561,322
;  CURRENT FILING DATE:  2000-04-28
;  NUMBER OF SEQ ID NOS:  12
; SEQ ID NO 2
;  LENGTH: 390
;  TYPE: PRT
;  ORGANISM: Ehrlichia Canis
;  FEATURE:
;  OTHER INFORMATION: P43 Antigen of Ehrlichia Canis
US-09-561-322-2

  Query Match        45.5%;  Score 1963;  DB 3;  Length 390;
  Best Local Similarity 99.7%;  Pred. No. 7.4e-161;
  Matches 373;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy      1 PSGDIQDSQQDQQEQDQQQGAVGGAVGNSPIERERVAAPESEDLYTVIIPKGKRTAAPI 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     16 PSGDIQDSQQDQQEQDQQQGAVGGAVGNSPIERERVAAPESEDLYTVIIPKGKRTAAPI 75

Qy     61 LERKSPTPEPKVEDDEDLPPTLPPRTFSGEGYDDVGVSMPTVSRGIYQPPIVQDSNLYSS 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 LERKSPTPEPKVEDDEDLPPTLPPRTFSGEGYDDVGVSMPTVSRGIYQPPIVQDSNLYSS 135

Qy    121 IGGVPQEAQYDAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDVRNAVLGNKEIKEE 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 IGGVPQEAQYDAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDVRNAVLGNKEIKEE 195

Qy    181 WLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVKG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 WLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVKG 255

Qy    241 AKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGD 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 AKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGD 315

Qy    301 KLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPL 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    316 KLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPL 375

Qy    361 HHAVEHLPPVILKG 374
            || ||||||| |||
```

```
Db   376 HHAVEHLPPVNLKG 389

RESULT 2
US-09-949-016-6966
; Sequence 6966, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6966
; LENGTH: 1719
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6966

Query Match             9.1%;  Score 390.5; DB 4;  Length 1719;
Best Local Similarity  26.0%;  Pred. No. 4.9e-24;
Matches  174; Conservative 100; Mismatches 246;  Indels 149; Gaps  31;

Qy  194 RIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVK--------GA----- 241
Db  256 RLLLDRGAQIETKTKDEL-----------------------TPLHCAARNGHVRISEILLDHGAPIQA 300
Qy  242 --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT 298
Db  301 KTKNGLSPIHMAAQGDHLD---CVRLLLQYDAEIDDI----------TLD---HLTPLHVAAHCG 349
Qy  299 GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK 357
Db  350 HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL 405
Qy  358 NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT-IAHYAVKNK 407
Db  406 TPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK 460
Qy  408 NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP 467
Db  461 ------------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP 506
Qy  468 LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK 525
Db  507 LHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA 565
Qy  526 GEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE 585
Db  566 GKN-----------GLTPLHVAVHHNNLDIVKL--------------------------- 587
Qy  586 LVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP 645
Db  588 LLPRG------GSP--------HSPAWN-GYTPLHIAAKQNQVEVARSLLQY-GGSANAE 631
Qy  646 SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG 705
Db  632 SVQGVTPLHLAAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHLVAQEGHVPVADVLI- 688
Qy  706 QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG 763
Db  689 KHGVMDVATTRMG---YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ 744
Qy  764 GDPKIVSSLLKAGAVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL--LDRGADINLD 820
```

```
Db  745 GHTDIVTLLLKNGASPNEVSSDGTTPLAIA-----KRLGYISVTDVLKVVTDETSFVLVS 799

Qy  821 GDHNILFDQ 829

Db  800 DKHRMSFPE 808


RESULT 3
US-09-949-016-6964
; Sequence 6964, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6964
; LENGTH: 1856
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6964

Query Match             9.1%;  Score 390.5; DB 4;  Length 1856;
Best Local Similarity  26.0%;  Pred. No. 5.6e-24;
Matches  174; Conservative 100; Mismatches 246;  Indels 149; Gaps  31;

Qy  194 RIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVK--------GA----- 241
Db  256 RLLLDRGAQIETKTKDEL-----------------------TPLHCAARNGHVRISEILLDHGAPIQA 300
Qy  242 --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT 298
Db  301 KTKNGLSPIHMAAQGDHLD---CVRLLLQYDAEIDDI----------TLD---HLTPLHVAAHCG 349
Qy  299 GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK 357
Db  350 HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL 405
Qy  358 NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT-IAHYAVKNK 407
Db  406 TPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK 460
Qy  408 NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP 467
Db  461 ------------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP 506
Qy  468 LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK 525
Db  507 LHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA 565
Qy  526 GEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE 585
Db  566 GKN-----------GLTPLHVAVHHNNLDIVKL--------------------------- 587
Qy  586 LVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLIHLP 645
Db  588 LLPRG------GSP--------HSPAWN-GYTPLHIAAKQNQVEVARSLLQY-GGSANAE 631
Qy  646 SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG 705
Db  632 SVQGVTPLHLAAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHLVAQEGHVPVADVLI- 688
```

Left column

```
Qy  706  QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG  763
Db  689  KHGVMVDATTRMG---YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ  744

Qy  764  GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGREKGVLGIVNKL---LDRGADINLD  820
Db  745  GHTDIVTLLLKNGASPNEVSSDGTTPLAIA----KRLGYISVTDVLKVVTDETSFVLVS  799

Qy  821  GDHNILFDQ  829
Db  800  DKHRMSFPE  808
```

RESULT 4
US-09-949-016-5876
; Sequence 5876, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5876
; LENGTH: 1880
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-5876

```
Query Match           9.1%;   Score 390.5;  DB 4;  Length 1880;
Best Local Similarity 26.0%;  Pred. No. 5.7e-24;
Matches 174;  Conservative 100;  Mismatches 246;  Indels 149;  Gaps 31;

Qy  194  RIVASKGDLSEDQVEBILDIIFMNESEIAEGISNPLHADVDNNPVK-------GA----  241
Db  256  RLLLDRGAQIETKTKDEL---------------TPLHCAARNGHVRISEILLDHGAPIQA  300

Qy  242  --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT  298
Db  301  KTKNGLSPIHMAAQGDHLD------TLD---HLTPLHVAAHCG                  349

Qy  299  GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK  357
Db  350  HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL  405

Qy  358  NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT-IAHYAVKNK  407
Db  406  TPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK  460

Qy  408  NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP  467
Db  461  ------------AKVNAKADDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP     506

Qy  468  LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK  525
Db  507  LHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA  565

Qy  526  GEDTVAKSLLMEDYKGFTPLHFVAGGSRDTFRVRKNYEKCHDLATIRAALMQDRSGGE   585
Db  566  ---------GLTPLHVAVHHNLDIVKL------------                      587

Qy  586  LVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP  645
Db  588  LLPRG------GSP------HSPAWN-GYTPLHLIAAKQNQVEVARSLLQY-GGSANAE   631
```

Right column

```
Qy  646  SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG  705
Db  632  SVQGVTPLHLAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHLVAQEGHVPVADVLI-   688

Qy  706  QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG  763
Db  689  KHGVMVDATTRMG---YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ  744

Qy  764  GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGREKGVLGIVNKL---LDRGADINLD  820
Db  745  GHTDIVTLLLKNGASPNEVSSDGTTPLAIA----KRLGYISVTDVLKVVTDETSFVLVS  799

Qy  821  GDHNILFDQ  829
Db  800  DKHRMSFPE  808
```

RESULT 5
US-09-949-016-6965
; Sequence 6965, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6965
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6965

```
Query Match           9.1%;   Score 390.5;  DB 4;  Length 1881;
Best Local Similarity 26.0%;  Pred. No. 5.7e-24;
Matches 174;  Conservative 100;  Mismatches 246;  Indels 149;  Gaps 31;

Qy  194  RIVASKGDLSEDQVEBILDIIFMNESEIAEGISNPLHADVDNNPVK-------GA----  241
Db  256  RLLLDRGAQIETKTKDEL---------------TPLHCAARNGHVRISEILLDHGAPIQA  300

Qy  242  --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT  298
Db  301  KTKNGLSPIHMAAQGDHLD------TLD---HLTPLHVAAHCG                  349

Qy  299  GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK  357
Db  350  HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL  405

Qy  358  NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT-IAHYAVKNK  407
Db  406  TPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK  460

Qy  408  NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP  467
Db  461  ------------AKVNAKADDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP     506

Qy  468  LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK  525
Db  507  LHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA  565

Qy  526  GEDTVAKSLLMEDYKGFTPLHFVAGGSRDTFRVRKNYEKCHDLATIRAALMQDRSGGE   585
```

```
Db  566 GKN---------------GLTPLHVAVHHNLDIVKL-------------------------- 587
Qy  586 LVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP 645
        |:|  |||       ||: :|||:    :|||||
Db  588 LLPRG---------GSP------HSPAWN-GYTPLHIAAKQNQVEVARSLLQY-GGSANAE 631
Qy  646 SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG 705
        :|  |:|||:|||       ||:|
Db  632 SVQGVTPLHLAAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHIVAQEGHVPVADVLI- 688
Qy  706 QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG 763
        ||  ||||:|| |    :|  |||::||||           :|| |:||
Db  689 KHGVMVDATTRMG---YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ 744
Qy  764 GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL---LDRGADINLD 820
        ||:| |||:|||||||||:::|||  |||: |   :    |::||||   ||||||||||
Db  745 GHTDIVTLLLKNGASPNEVSSDGTTPLAIA------KRLGYISVTDVLKVVTDETSFVLVS 799
Qy  821 GDHNILFDQ 829
        |:|:|
Db  800 DKHRMSFPE 808
```

RESULT 6
US-09-949-016-9010
; Sequence 9010, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9010
; LENGTH: 1883;
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9010

Query Match   9.1%; Score 390.5; DB 4; Length 1883;
Best Local Similarity   26.0%; Pred. No. 5.7e-24;
Matches 174; Conservative 100; Mismatches 246; Indels 149; Gaps 31;

```
Qy  194 RIVASKGDLSEDQVEILDIIFMNESEIAEGISNPLHADVDNNPVK--------GA---- 241
        |::|:|:|                                 :TPLHCAARNGHVRISEILLDHGAPIQA 303
Db  259 RLLLDRGAQIETKTKDEL-----------------------
Qy  242 --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT 298
        :| :: |                        TLD---HLTPLHVAAHCG 352
Db  304 KTKNGLSPIHMAAQGDHLD--CVRLLLQYDAEIDDI------
Qy  299 GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK 357
        |:|:||
Db  353 HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL 408
Qy  358 NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT-IAHYAVKNK 407
        |:|  ||       :|:|                          VETPLHMAARAGHTEVAKYLLQNK 463
Db  409 TPLHVASFMGHLPIVKNLLQRGASPNVSNVK------
Qy  408 NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP 467
Db  464 ----------AKVNAKAKDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP 509
Qy  468 LHIAVEBGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK 525
```

RESULT 7
US-09-949-016-9011
; Sequence 9011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9011
; LENGTH: 1883;
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9011

Query Match   9.1%; Score 390.5; DB 4; Length 1883;
Best Local Similarity   26.0%; Pred. No. 5.7e-24;
Matches 174; Conservative 100; Mismatches 246; Indels 149; Gaps 31;

```
Qy  194 RIVASKGDLSEDQVEILDIIFMNESEIAEGISNPLHADVDNNPVK--------GA---- 241
        |::|:|:|                                 :TPLHCAARNGHVRISEILLDHGAPIQA 303
Db  259 RLLLDRGAQIETKTKDEL-----------------------
Qy  242 --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT 298
        :| :: |                        TLD---HLTPLHVAAHCG 352
Db  304 KTKNGLSPIHMAAQGDHLD--CVRLLLQYDAEIDDI------
Qy  299 GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK 357
        |:|:||
Db  353 HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL 408
Qy  358 NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT-IAHYAVKNK 407
Db  409 TPLHVASFMGHLPIVKNLLQRGASPNVSNVK------VETPLHMAARAGHTEVAKYLLQNK 463
```

```
Db  510 LHIAAREGHEVTVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRAELLLERDAHPNAA 568
Qy  526 GEDTVAKSLLMEDYKGFTPLHFVAGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE 585
        |:|:       ||||        :: :
Db  569 GKN-----------GLTPLHVAVHHNNLDIVKL------------------------- 590
Qy  586 LVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP 645
        |:|  |||       ||: :|||:    :|||||
Db  591 LLPRG---------GSP------HSPAWN-GYTPLHIAAKQNQVEVARSLLQY-GGSANAE 634
Qy  646 SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG 705
        :|  |:|||:|||       ||:|
Db  635 SVQGVTPLHLAAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHIVAQEGHVPVADVLI- 691
Qy  706 QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG 763
        ||  ||||:|| |    :|  |||::||||           :|| |:||
Db  692 KHGVMVDATTRMG---YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ 747
Qy  764 GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL---LDRGADINLD 820
        ||:| |||:|||||||||:::|||  |||: |   :    |::||||   ||||||||||
Db  748 GHTDIVTLLLKNGASPNEVSSDGTTPLAIA------KRLGYISVTDVLKVVTDETSFVLVS 802
Qy  821 GDHNILFDQ 829
        |:|:|
Db  803 DKHRMSFPE 811
```

Qy   408   NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP   467

Db   464   ------------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP   509

Qy   468   LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK   525

Db   510   LHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA   568

Qy   526   GEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE   585

Db   569   GKN------------GLTPLHVAVHHNNLDIVKL------------------------   590

Qy   586   LVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP   645

Db   591   LLPRG------GSP------HSPAWN-GYTPLHIAAKQNQVEVARSLLQY-GGSANAE   634

Qy   646   SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG   705

Db   635   SVQGVTPLHLAAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHLVAQEGHVPVADVLI-   691

Qy   706   QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG   763

Db   692   KHGVMVDATTRMG---YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ   747

Qy   764   GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL---LDRGADINLD   820

Db   748   GHTDIVTLLLKNGASPNEVSSDGTTPLAIA-----KRLGYISVTDVLKVVTDETSFVLVS   802

Qy   821   GDHNILFDQ   829

Db   803   DKHRMSFPE   811

RESULT 8
US-09-949-016-9012
; Sequence 9012, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9012
; LENGTH: 1883
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9012

Query Match   9.1%;   Score 390.5;   DB 4;   Length 1883;
Best Local Similarity   26.0%;   Pred. No. 5.7e-24;
Matches   174;   Conservative   100;   Mismatches 246;   Indels 149;   Gaps   31;

Qy   194   RIVASKGDLSEDQVEBILDIIFMNESEIAEGISNPLHADVDNNPVK-------GA----   241

Db   259   RLLLDRGAQIETKTKDEL---------------TPLHCAARNGHVRISEILLDHGAPIQA   303

Qy   242   --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT   298

Qy   304   KTKNGLSPIHMAAQGDHLD--CVRLLLQYDAEIDDI-----TLD---HLTPLHVAAHCG   352

Qy   299   GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK   357

Db   353   HHRVAKVLLDKGAKPNSRALN--GFFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL   408

RESULT 9
US-09-949-016-9013
; Sequence 9013, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9013
; LENGTH: 1883
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9013

Query Match   9.1%;   Score 390.5;   DB 4;   Length 1883;
Best Local Similarity   26.0%;   Pred. No. 5.7e-24;
Matches   174;   Conservative   100;   Mismatches 246;   Indels 149;   Gaps   31;

Qy   194   RIVASKGDLSEDQVEBILDIIFMNESEIAEGISNPLHADVDNNPVK-------GA----   241

Db   259   RLLLDRGAQIETKTKDEL---------------TPLHCAARNGHVRISEILLDHGAPIQA   303

Qy   242   --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT   298

```
Db    304  KTKNGLSPIHMAAQGDHLD--CVRLLLQYDAEIDDI------TLD---HLTPLHVAAHCG  352
                ||||||||        |:::: ||      :::                |||||||||
Qy    299  GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK  357

Db    353  HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL  408
                :|||||||        | |:::  |||||||||||||||||||        ::
Qy    358  NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF---GNT-IAHYAVKNK  407

Db    409  TPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK  463
                |||||||:        ::::::::       ||::::        |||||||
Qy    408  NADLTLFNMLKASGADLVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP  467

Db    464  ------------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP  509
                            |||||||||        |:|||||||        :|||||
Qy    468  LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK  525

Db    510  LHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA  568
                |||||        |||||        ||||||||||||||||        ||
Qy    526  GEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE  585

Db    569  GKN------GLTPLHVAVHHNNLDIVKL-  590
                              |||||||||
Qy    586  LVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP  645

Db    591  LLPRG--------GSP------HSPAWN-GYTPLHIAAKQNQEVARSLLQY-GGSANAE  634
                |||        |||        |||||||||||||||||
Qy    646  SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG  705

Db    635  SVQGVTPLHLAAQEGHAEMVALLL---SKQANGNLGNKSGLTPLHLVAQEGHVPVADVLI-  691
                ::||||||||        |:|        :||||||        |||||||
Qy    706  QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG  763

Db    692  KHGVMVDATTRMG---YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ  747
                |||:::        ||||||||||        :|||||||
Qy    764  GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL---LDRGADINLD  820

Db    748  GHTDIVTLLLKNGASPNEVSSDGTTPLAIA------KRLGYISVTDVLKVVTDETSFVLVS  802
                :::||||||||        :::::::                |||||||||||||||||
Qy    821  GDHNILFDQ  829

Db    803  DKHRMSFPE  811
```

RESULT 10
US-09-949-016-9014
; Sequence 9014, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9014
;   LENGTH: 1883
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-9014

```
Query Match         9.1%;  Score 390.5;  DB 4;  Length 1883;
Best Local Similarity   26.0%;  Pred. No. 5.7e-24;
Matches  174;  Conservative  100;  Mismatches  246;  Indels  149;  Gaps  31;

Qy    194  RIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVK-------GA----  241
```

RESULT 11
US-09-949-016-9015
; Sequence 9015, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig et al.
;   TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;   TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;   FILE REFERENCE: CL001307
;   CURRENT APPLICATION NUMBER: US/09/949,016
;   CURRENT FILING DATE: 2000-04-14
;   PRIOR APPLICATION NUMBER: 60/241,755
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 60/237,768
;   PRIOR FILING DATE: 2000-10-03
;   PRIOR APPLICATION NUMBER: 60/231,498
;   PRIOR FILING DATE: 2000-09-08
;   NUMBER OF SEQ ID NOS: 207012
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9015
;   LENGTH: 1883
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-9015

```
Query Match              9.1%;  Score 390.5;  DB 4;  Length 1883;
Best Local Similarity  26.0%;  Pred. No. 5.7e-24;
Matches  174;  Conservative  100;  Mismatches  246;  Indels  149;  Gaps  31;

Qy    194  RIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVK------GA----- 241
           ::: :|: ::|::  |:::|  : :|: :: |:|
Db    259  RLLLDRGAQIETKTKDEL----------------------------TPLHCAARNGHVRISEILLDHGAPIQA 303

Qy    242  --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT 298
             :::|:.|:|   |||  :|
Db    304  KTKNGLSPIHMAAQGDHLD--CVRLLLQYDAEIDDI------TLD---HLTPLHVAAHCG 352

Qy    299  GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK 357
           ::|: :|: :|     |||| ::::
Db    353  HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRMELLLKTGASID--AVTESGL 408

Qy    358  NPLHHA---VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT-IAHYAVKNK 407
           |||        ||||    :    |::|: :    |
Db    409  TPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK 463

Qy    408  NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP 467
           : |  |: :                ||| : |:||: :
Db    464  ------------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP 509

Qy    468  LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK 525
           ||||     |  |:::||:|||     ||:|||:|:||:|||||||:
Db    510  LHIAAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA 568

Qy    526  GEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE 585
           :    |       |  ||||||||
Db    569  GKN----------GLTPLHVAVHHNNLDIVKL- 590

Qy    586  LVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP 645
           |:    ||     |||:|     :|||      |||
Db    591  LLPRG------GSP------HSPAWN-GYTPLHIIAAKQNQVEVARSLLQY-GGSANAE 634

Qy    646  SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG 705
           | |:|| :||| |||     :||    ::|    |:|
Db    635  SVQGVTPLHLAAQEGHAEMVALLL---SKQANGNLGNKSGLTPLHLVAQEGHVPVADVLI- 691

Qy    706  QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG 763
           :      ||:|::|     ||||: ||:   |:|    :|||
Db    692  KHGVMVDATTRMG---YTPLHVASHYGNIKLVKFLLQHG-ADVNAKTKLGYSPLHQAAQQ 747

Qy    764  GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL---LDRGADINLD 820
           ||| :|||||:||:||  ||:|||:|  :: ||
Db    748  GHTDIVTLLLKNGASPNEVSSDGTTPLAIA------KRLGYISVTDVLKVVTDETSFVLVS 802

Qy    821  GDHNILFDQ 829
           |||: :|:
Db    803  DKHRMSFPE 811
```

RESULT 12
US-09-949-016-9016
; Sequence 9016, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9016

; LENGTH: 1883
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9016

```
Query Match              9.1%;  Score 390.5;  DB 4;  Length 1883;
Best Local Similarity  26.0%;  Pred. No. 5.7e-24;
Matches  174;  Conservative  100;  Mismatches  246;  Indels  149;  Gaps  31;

Qy    194  RIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVK------GA----- 241
           ::: :|: ::|::  |:::|  : :|: :: |:|
Db    259  RLLLDRGAQIETKTKDEL----------------------------TPLHCAARNGHVRISEILLDHGAPIQA 303

Qy    242  --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT 298
             :::|:.|:|   |||  :|
Db    304  KTKNGLSPIHMAAQGDHID--CVRLLLQYDAEIDDI------TLD---HLTPLHVAAHCG 352

Qy    299  GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK 357
           ::|: :|: :|     |||| ::::
Db    353  HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRMELLLKTGASID--AVTESGL 408

Qy    358  NPLHHA---VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT-IAHYAVKNK 407
           |||        ||||    :    |::|: :    |
Db    409  TPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK 463

Qy    408  NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP 467
           : |  |: :                ||| : |:||: :
Db    464  ------------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP 509

Qy    468  LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK 525
           ||||     |  |:::||:|||     ||:|||:|:||:|||||||:
Db    510  LHIAAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA 568

Qy    526  GEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE 585
           :    |       |  ||||||||
Db    569  GKN----------GLTPLHVAVHHNNLDIVKL- 590

Qy    586  LVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP 645
           |:    ||     |||:|     :|||      |||
Db    591  LLPRG------GSP------HSPAWN-GYTPLHIIAAKQNQVEVARSLLQY-GGSANAE 634

Qy    646  SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG 705
           | |:|| :||| |||     :||    ::|    |:|
Db    635  SVQGVTPLHLAAQEGHAEMVALLL---SKQANGNLGNKSGLTPLHLVAQEGHVPVADVLI- 691

Qy    706  QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG 763
           :      ||:|::|     ||||: ||:   |:|    :|||
Db    692  KHGVMVDATTRMG---YTPLHVASHYGNIKLVKFLLQHG-ADVNAKTKLGYSPLHQAAQQ 747

Qy    764  GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL---LDRGADINLD 820
           ||| :|||||:||:||  ||:|||:|  :: ||
Db    748  GHTDIVTLLLKNGASPNEVSSDGTTPLAIA------KRLGYISVTDVLKVVTDETSFVLVS 802

Qy    821  GDHNILFDQ 829
           |||: :|:
Db    803  DKHRMSFPE 811
```

RESULT 13
US-09-949-016-9017
; Sequence 9017, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9017
; LENGTH: 1883
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9017

Query Match         9.1%;  Score 390.5; DB 4; Length 1883;
Best Local Similarity   26.0%;  Pred. No. 5.7e-24;
Matches 174; Conservative 100; Mismatches 246; Indels 149; Gaps 31;

Qy  194 RIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVK------GA----  241
Db  259 RLLLDRGAQIETKTKDEL-----------TPLHCAARNGHVRISEILLDHGAPIQA    303

Qy  242 --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT  298
Db  304 KTKNGLSPIHMAAQGDHLD--CVRILLQYDAEIDDI------TLD--HLTPLHVAAHCG    352

Qy  299 GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK  357
Db  353 HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKQNHVRVMELLLKTGASID--AVTESGL  408

Qy  358 NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF----GNT-IAHYAVKNK  407
Db  409 TPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK  463

Qy  408 NADLTLFNMLKASGADLNVRNVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP   467
Db  464 -------------AKVNAKAKDDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP   509

Qy  468 LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK  525
Db  510 LHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA  568

Qy  526 GEDTVAKSILLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE 585
Db  569 GKN------------------GLTPLHVAVHHNNLDIVKL----------------     590

Qy  586 LVNLGDFESENILGGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP 645
Db  591 LLPRG------GSP------HSPAWN-GYTPLHLIAAKQNQVEVARSLLQY--GGSANAE   634

Qy  646 SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG  705
Db  635 SVQGVTPLHLAAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHLVAQEGHVPVADVLI-  691

Qy  706 QHG---IDINQRMGSDKNTVLHYAISKGDSFIVQKILAHTGVDVNCENNLGQTPLHLAVEG 763
Db  692 KHGWVDATTRMG---YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ   747

Qy  764 GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL------LDRGADINLD 820
Db  748 GHTDIVTLLLKNGASPNEVSSDGTTPLAIA------KRLGYISVTDVLKVVTDETSFVLVS  802

Qy  821 GDHNILFDQ 829
Db  803 DKHRMSFPE 811

RESULT 14
US-09-949-016-6978
; Sequence 6978, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6978
; LENGTH: 4377
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6978

Query Match         8.8%;  Score 377.5;  DB 4;  Length 4377;
Best Local Similarity   25.5%;  Pred. No. 3e-22;
Matches 175; Conservative 92; Mismatches 277; Indels 141; Gaps 26;

Qy  184 TSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNES-----EIAEGISNPLH------   230
Db  197 TKGKVRLPALHIAARKDD--------TKAAALLQNDNNADVESKSGFT-PLHIAAHYGNIN 249

Qy  231 -------ADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNV   281
Db  250 VATLLLNRAAAVDFT--ARNDITPLH-VASKRGNANMVKLL------LDRGAK-IDA    296

Qy  282 LDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIE   341
Db  297 KTRDGLTPLHCGARSGHEQVEMLLDRAAPILSKTKN--GLSPLHMATQGDHLNCVQL-   352

Qy  342 FTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAH  401
Db  353 LLQHNVPVDDVTNDYLTALHVAAHCGHYKVAKVLLDKKANP------NAKALNGFTPLH   405

Qy  402 YAVKNKNADLTLFNMLKASGADLNVRNVGRAPIHVASSNGKANAVSGLVSCGIDVNSQD   461
Db  406 IACK-KNR-IKVMELLLKHGASIQAVTESGLTPIHVAAFMGHVNIVSQLMHHGASPNTTN  463

Qy  462 VNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GS 519
Db  464 VRGETALHMAARSGQAE-VVRYLVQDGAQVEAKAKDDQTPLHISARLGKADIVQQLLQQG  522

Qy  520 AKPNI--------------------KSLLMEDYKGFTPLHFVAGGGSRDTFR         558
Db  523 ASPNAATTSCYTPLHLSAREGHEDVAAFLLDHGASLSITTKKGFTPLHVAA-------   573

Qy  559 VVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENLI-----GSPNAKFLQHIQSANF 614
Db  574 -------------------------------------------------KYGKLEVANLLLQKSASPDA- 597

Qy  615 GFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAKSAKP  674
Db  598 GLTPLHVAAHYDNQKVALLLLD--QCASPHAAAKNGYTPLHIAAKKNQMDIATTLLEYG-- 654

Qy  675 SDLNFKTSATPTPLNLACLRGDNEVVRGLVGQH-GIDINQRMGSDKNTVLHYAISKGDSF  733
Db  655 ADANAVTRQGIASVHLAAQEGHVDMVSLLLGRNANVNLSNKSGL---TPLHLA-AQEDRV  710

Qy  734 LVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRLDDNGRSVLSSA  793
Db  711 NVAEVLVNQGAHVDAQTKMGYTPLHVGCHYGNIKIVNFLLQHSAKVNAKTKNGYTPLHQA  770

Qy  794 IVPGRKEKGVLGIVNKLLDRGADIN 818
Db  771 ----AQQGHTHINVLLQNNASPN 790

RESULT 15
US-10-164-595-38
; Sequence 38, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
```

Search completed: February 26, 2005, 18:55:45
Job time : 50 secs

```
; APPLICANT: OriGene Technologies, Inc
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
; FILE REFERENCE: 1U 103 R1
; CURRENT APPLICATION NUMBER: US/10/164,595
; CURRENT FILING DATE:  2002-06-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 38
;   LENGTH: 994
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-164-595-38

Query Match         8.6%;  Score 370.5;  DB 4;  Length 994;
Best Local Similarity  22.9%;  Pred. No. 1.1e-22;
Matches  191;  Conservative  114;  Mismatches  311;  Indels  217;  Gaps   28;

Qy    108 QPPIVGDSNLYSSIGGVPQEAQYDAAARAGGPRKFLYGPYTFSNGQEIMDEFDTPWPDV 167
              ||::||         || | : ::                     :: : ||: ||
Db      9 QPPLVQ------AIFSGDPEEI----------RMLIHK----TEDVNTLDSEKRTPLHVA 48

Qy    168 RNAVLGNKEIKE----------------EWLTTSGPVRDIADRIVASKGDLSEDQVEEILDI 213
            || :||:||         | ||||   :  ||||  :  ::
Db     49 --AFLGDAEIIELLILSGARVNAKDNWLT---PLH----RAVASRS-------EEAVQV 92

Qy    214 IFMNESEIAEGISN---PLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVEND 270
            : :::|||            |::|   | : |||
Db     93 LIKHSADVNARDKNWQTPLHVAAANKAVKCABVIIPLL------------------- 130

Qy    271 EGDLGANAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQ 330
                         ||||||:  :|:|  :  |: : |: :
Db    131 -----SSVNVSDRGGRTALHHAALNGHVEMVNLLLAKGAN--INAFDKKDRRALHWAAY 182

Qy    331 KPDCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVN 390
            : :| |              |   ||  :   :::  ||:| :||
Db    183 MGHLDVVAL--LINHGAEVTCKDKKGYTPLHAAASN-------GQINVVKHLLNLGVEID 233

Qy    391 DPDYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIH-VASSNGKANAVSG 449
            : :|||         :|:|  :|:|| |:||||| : | |   |: ||| | |  ||
Db    234 EINVYGNTALHIAC--YNGQDAVVNELIDYGANVNQPNNNGFTPLHFAAAASTHGALCLEL 291

Qy    450 LVSCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYG 509
            ||:           | ||||    | :::  | |:  :||| : |:: | |  ||| ||
Db    292 LVNNGADVNIQSKDGKSPLHMTAVHGRP-TRSQTLIQNGGEIDCVDKDGNTPLHVAARYG 350

Qy    510 DIGVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKC-- 567
            ||:  | :              :|::        : :  || | |  | : | :
Db    351 HELLINTLITS-----GADTAKCGI----HSMFPLHLAALNAHSDCCRKLLSSGQKYSI 400

Qy    568 ------HDLATIRAALMQDRSGGGELVNL----GDFESENILGSPNAKFLQ--------- 607
                     | | | : ||  :::|    ||: |: ||
Db    401 VSLFSNEHVLSAGFEIDTPDKFGRTCLHAAAAAGGNVECIKLLQSSGADFHKKDKCCGRTPL 460

Qy    608 HIQSANFGP--------------------------------SPARRGIVSSNHNVMKDI-- 634
            |:| |||                                  |   |:|:|| | :::
Db    461 HYAAANCHFHCIETLVTTGANVNETDDWGRTALHYAAASDMDRNKTILGNAHDNSEELER 520

Qy    635 ----------LNFVGDSLHLPSER---GYNAMQVAALFGDKEAVRMLAKSAKPS---- 675
                      |: |||| || :|   |:| : ||:  | :::::
Db    521 ARELKEKEATLCLEFLLQNDANPSIRDKEGYNSIHYAAAYGHRQCLELLLERTNSGFEES 580

Qy    676 -----------------------DLNFKTSATPTPLNLACLRGDNEVVRGLVGQ 706
                                   |:| ||::| |:| |||:|:||  :|| |
Db    581 DSGATKSPLHLAAYNGHHQALEVLLQSLVDLLDIRDEKGRTALDLAAFKGHTECVEALINQ 640

Qy    707 HGIDINQRMGGSDKNTVLHYAISKGDSFLVQKIL--AHTGVDVNCENNLGQTPLHLAVEGG 764
            ||||:::: ||| |:|  |: | :||    |:|   | |||||:  | |:|||||||| |
Db    641 -GASIFVKDNVTKRTPLHASVINGHTLCLRLLLEIADNPEAVDVKDAKGQTPLMLAVAYG 699

Qy    765 DPKIVSSLLKAGAVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADI 817
            | |::| :|:|  |||   ::|  | :      :| |:   |  |:| ||:
Db    700 HIDAVSLLLEKEANVDTVDIIGCTALHRGIMTGHEE-----CVQMLLEQEVSI 747
```

THIS PAGE BLANK (U)

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        February 26, 2005, 18:54:00 ; Search time 143 Seconds
                                             (without alignments)
                                   1906.329 Million cell updates/sec

Title:            US-10-701-038-2
Perfect score:    4310
Sequence:         1 PSGDIQDSQQDQQQEQDQQQ.........DRGADINLDGDHNILFDQCL 831

Scoring table:    BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched:         138339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters:   1385339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :     Published Applications AA:*
     1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
     2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
     3:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
     4:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
     5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
     6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
     7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
     8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
     9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
    10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
    11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
    12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
    13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
    14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
    15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
    16:  /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
    17:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
    18:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
    19:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
    20:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	4310	100.0	831	16	US-10-701-038-2	Sequence 2, Appli
2	1300.5	30.2	831	16	US-10-701-038-1	Sequence 1, Appli
3	421	9.8	1053	15	US-10-291-172-343	Sequence 343, App
4	421	9.8	1053	15	US-10-221-278-343	Sequence 343, App
5	373.5	8.7	1056	17	US-10-479-764-18	Sequence 18, Appl
6	368.5	8.5	1330	15	US-10-108-260A-3237	Sequence 3237, Ap
7	368	8.5	1619	14	US-10-369-978-4	Sequence 4, Appli
8	364.5	8.5	919	15	US-10-108-260A-4122	Sequence 4122, Ap
9	351	8.1	1709	14	US-10-369-978-6	Sequence 6, Appli
10	350	8.1	1094	17	US-10-479-764-22	Sequence 22, Appl
11	345.5	8.0	1762	14	US-10-205-194-117	Sequence 117, App
12	344	8.0	720	15	US-10-433-794-20	Sequence 20, Appl
13	343	8.0	1724	9	US-09-964-899-43	Sequence 43, Appl
14	327	7.6	765	15	US-10-182-243-56	Sequence 56, Appl
15	326.5	7.6	1250	17	US-10-479-764-1	Sequence 1, Appli
16	324.5	7.5	765	14	US-10-128-174-3	Sequence 3, Appli
17	324.5	7.5	765	14	US-10-128-174-34	Sequence 34, Appl
18	324.5	7.5	765	14	US-10-128-174-35	Sequence 35, Appl
19	324.5	7.5	765	14	US-10-128-174-36	Sequence 36, Appl
20	324.5	7.5	765	14	US-10-128-174-37	Sequence 37, Appl
21	324.5	7.5	765	14	US-10-128-174-38	Sequence 38, Appl
22	324.5	7.5	765	14	US-10-128-174-39	Sequence 39, Appl
23	324.5	7.5	765	14	US-10-128-174-40	Sequence 40, Appl
24	324.5	7.5	765	14	US-10-128-174-41	Sequence 41, Appl
25	324.5	7.5	765	14	US-10-128-174-42	Sequence 42, Appl
26	324.5	7.5	765	14	US-10-128-174-43	Sequence 43, Appl
27	324.5	7.5	765	14	US-10-128-174-44	Sequence 44, Appl
28	324.5	7.5	3913	15	US-10-334-143-45	Sequence 45, Appl
29	324	7.5	1704	14	US-10-369-978-2	Sequence 2, Appli
30	320	7.4	949	9	US-09-841-835-10	Sequence 10, Appl
31	320	7.4	1327	9	US-09-841-835-2	Sequence 2, Appli
32	320	7.4	1327	10	US-09-972-115A-8	Sequence 8, Appli
33	320	7.4	1327	14	US-10-199-937-4	Sequence 4, Appli
34	314.5	7.3	602	15	US-10-648-512-28	Sequence 28, Appl
35	314.5	7.3	898	15	US-10-648-512-24	Sequence 24, Appl
36	314.5	7.3	906	15	US-10-648-512-36	Sequence 36, Appl
37	314.5	7.3	906	15	US-10-648-512-38	Sequence 38, Appl
38	314.5	7.3	970	15	US-10-648-512-34	Sequence 34, Appl
39	314.5	7.3	1001	15	US-10-648-512-40	Sequence 40, Appl
40	314.5	7.3	1065	15	US-10-648-512-22	Sequence 22, Appl
41	313.5	7.3	1065	15	US-10-648-512-32	Sequence 32, Appl
42	308.5	7.2	1065	15	US-10-648-512-42	Sequence 42, Appl
43	307.5	7.1	1119	16	US-10-757-262-76	Sequence 76, Appl
44	304	7.1	1050	15	US-10-250-613-3	Sequence 3, Appli
45	302.5	7.0	510	15	US-10-648-512-26	Sequence 26, Appl

ALIGNMENTS

```
RESULT 1
US-10-701-038-2

; Sequence 2, Application US/10701038
; Publication No. US20040121433A1
; GENERAL INFORMATION:
;  APPLICANT:  McBride, Jere W.
;  APPLICANT:  Walker, David H.
;  TITLE OF INVENTION:  P153 and P156 Antigens for the Immunodiagnosis of
;  TITLE OF INVENTION:  Canine and Human Ehrlichioses and Uses Thereof
;  FILE REFERENCE:  D6481
;  CURRENT APPLICATION NUMBER:  US/10/701,038
;  CURRENT FILING DATE:  2003-11-04
;  PRIOR APPLICATION NUMBER:  US 60/423,573
;  PRIOR FILING DATE:  2002-11-04
;  NUMBER OF SEQ ID NOS:  2
;  SEQ ID NO 2
;  LENGTH:  831
;  TYPE:  PRT
;  ORGANISM:  Ehrlichia chaffeensis
;  FEATURE:
;  OTHER INFORMATION:  immunoreactive surface protein p156
US-10-701-038-2

  Query Match             100.0%;  Score 4310;  DB 16;  Length 831;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-298;
  Matches  831;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1  PSGDIQDSQQDQQQEQDQQQGAVGGAVGNSPIERERVAAPESEDLYTVIIPKGKRTAAPI  60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  PSGDIQDSQQDQQQEQDQQQGAVGGAVGNSPIERERVAAPESEDLYTVIIPKGKRTAAPI  60

Qy     61  LERKSPTPEPKVEDDEDLPPTLPPRTFSGEGYDDVGVSMPTVSRGIYQPPIVQDSNLYSS  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  LERKSPTPEPKVEDDEDLPPTLPPRTFSGEGYDDVGVSMPTVSRGIYQPPIVQDSNLYSS  120

Qy    121  IGGVPQEAQYDAAARAGGPRKFLYGPVTFSNGQEIMDFBFDTPWPDVRNAVLGNKEIKEE  180
```

Matches 297; Conservative 156; Mismatches 265; Indels 113; Gaps 16;

```
Qy   39  APSEDLITYTIIPKGKRTA-APILERKSPTPEPKVEDDEDLPPTLPPRTFSGE-------  90
            :  ::| :  ||: ||| .     ::|  :      ||::||||::
Db   11  AVEEDPLYAVPLPKGQRPAPTQVLEE-----DPSVEEEEIAPPLPPRRNVGEVBPQEDP  65

Qy   91  ------GY-----DDVGVSMPTVSR----GIYQPPIVQDS------NLY---SSIGGVP 125
               |:       :|        ::|   ||  ::|
Db   66  IYQGIPGHQEEMEEDPYASLDQVSQGAGADGIQENPVPQEAGELEEDIYQDPADFQGLG 125

Qy  126  QEA-QYDAAARAG---GPRKFLYGPYTFSNGQBIMDFEFDTPWPDVRNAVLGNKEIKEEW 181
          :   ::  ::|     | :  ||:|:|:| |  : ||:  :||::  :|::::|||:
Db  126  QGGQLDQAGYQGPSIGDRQLVNGPYGFNDGSYAMEFD-DVMWEGVRDAVIHDEIDPKF  184

Qy  182  LTTSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVKGA 241
          :| |::|  :  |:|: ::| | :::||:  |: ::|: :| |:: ||||::|:|||::
Db  185  LVTDGLMRHICDKIVQSEGNLPEPDLEEIVSILKNDKEGISELINEPVQVDIPNNPVREG 244

Qy  242  KNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGDK 301
          :|||||:||:  : :|||: :||  ::: ||||: ||: ||| ||||||||||||||
Db  245  RNVMTLLHLLAYAYNVDPRIINAIEBSVENSFGESGLDGYNIQDADGNLPLHHAAKNCNGQV 304

Qy  302  LKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVD-GDGKNPL 360
           ::|  ::.::: : :  :.|||||::  :|  || :|: |::.  :::. ||
Db  305  LDNCISKTNSNIININRNFGNQSPLHVMVQNPGCSIGNIQVANECGMDFNLIDHPTGRMPI 364

Qy  361  HHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAHYAVKNKNADLTLFNMLKAS 420
          |:|:|
Db  365  HYAAE-------AASSE-------VLSYVIRNTKAE------ 386

Qy  421  GADLNRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETV  480
            :                            :   ||:||::||||::||:|::.:.|
Db  387  ----------------SPQASA------VNTQDVNGRTPLHCAAISGNSKGL  416

Qy  481  LAVLNQRGADVSVQNNDGVTPMLSAAKYDIGVIKALGSAKPNIKGEDTVAKSLLMEDYK 540
           : :|:| | . |:|:| |::|:|::  ::  :|  ||:| : :.||::::  | :|::
Db  417  SVMLLQNGVDCAVRDKNYSTPLHYAVAGNDIKSIKNLCSVKGRVGVKSSAASLLCEDLQ 476

Qy  541  GFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGS 600
          | |||| .|:  :. :  .||:  ::  .||::|: :: .|. |  ::. ||:.:.|| |
Db  477  GDTPLHIACKVEGTKAFETVRQSIKKGHHGKQVLQELLIREGSGPRL-NVSGFGSQSILSG 535

Qy  601  PNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDIINFVGDSLHLPSERGYNAMQVAALFG 660
            :  : : :.||| .|||   : ::|:||:| |::|| :: : | ::|: | |:||||
Db  536  VSGDLYGYLNSQNFPTSPVHAAVKANNLQLINLFLKKSPDILRQSSPNGFNPVHMAALFA 595

Qy  661  DKEAVKMLAKSAKPSDLNFKTSATPTPLNALCLRGDNEVVRGLVGQHGIDINQRMGSDKN 720
          | ::||:: :.:   :.:   :  :|||:|:|||| .:::|:|  .:.:::: ||||.:|
Db  596  DVKTVKLIIENASGEEVNAQSDSTLTPLHLACIRGDGSIIKRMVEHESVNVNQTMGPDQN 655

Qy  721  TVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVN 780
          |||:||| :|:. |::::| | . |:|  .||:.|||::  :.|:|:. ||::..||.||
Db  656  TVLQYAINRGNHSLIKRLLSHPSIDLNVRNADGKTSAHSAMEKGDLKTVKALCNAGADVN 715

Qy  781  RLDDNGRSVLSSAIVPGREKEGKVLGIVNKLLDRGADINLDGDHNILFDQCL 831
           :|:|||||:|||| .| ::|:| | |||||||:.|| ::. |:||||.:|:
Db  716  TVDNNGRSVISSAIYSGGQNEKKLVPIVKLLLNSGAKIGSQEDKNILLQKCI 766
```

RESULT 3
US-10-291-172-343
; Sequence 343, Application US/10291172
; Publication No. US20030228584A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: No. US20030228584A1e1 Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-045
; CURRENT APPLICATION NUMBER: US/10/291,172
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 09/693,267
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/665,363
; PRIOR FILING DATE: 2000-09-19

```
Db  121  IGGVPQEAQYDAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDVRNAVLGNKEIKEE 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  181  WLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVKG 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  WLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVKG 240

Qy  241  AKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGD 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  AKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGD 300

Qy  301  KLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPL 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  KLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPL 360

Qy  361  HHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAHYAVKNKNADLTLFNMLKAS 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  HHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAHYAVKNKNADLTLFNMLKAS 420

Qy  421  GADLNRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETV  480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  GADLNRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETV  480

Qy  481  LAVLNQRGADVSVQNNDGVTPMLSAAKYDIGVIKALGSAKPNIKGEDTVAKSLLMEDYK 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  LAVLNQRGADVSVQNNDGVTPMLSAAKYDIGVIKALGSAKPNIKGEDTVAKSLLMEDYK 540

Qy  541  GFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGS 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  GFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGS 600

Qy  601  PNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDIINFVGDSLHLPSERGYNAMQVAALFG 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  PNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDIINFVGDSLHLPSERGYNAMQVAALFG 660

Qy  661  DKEAVKMLAKSAKPSDLNFKTSATPTPLNALCLRGDNEVVRGLVGQHGIDINQRMGSDKN 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661  DKEAVKMLAKSAKPSDLNFKTSATPTPLNALCLRGDNEVVRGLVGQHGIDINQRMGSDKN 720

Qy  721  TVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVN 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721  TVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVN 780

Qy  781  RLDDNGRSVLSSAIVPGREKEGKVLGIVNKLLDRGADINLDGDHNILFDQCL 831
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781  RLDDNGRSVLSSAIVPGREKEGKVLGIVNKLLDRGADINLDGDHNILFDQCL 831
```

RESULT 2
US-10-701-038-1
; Sequence 1, Application US/10701038
; Publication No. US20040121433A1
; GENERAL INFORMATION:
; APPLICANT: McBride, Jere W.
; APPLICANT: Walker, David H.
; TITLE OF INVENTION: P153 and P156 Antigens for the Immunodiagnosis of
; TITLE OF INVENTION: Canine and Human Ehrlichioses and Uses Thereof
; FILE REFERENCE: D6481
; CURRENT APPLICATION NUMBER: US/10/701,038
; CURRENT FILING DATE: 2003-11-04
; PRIOR APPLICATION NUMBER: US 60/423,573
; PRIOR FILING DATE: 2002-11-04
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: immunoreactive surface protein p153
US-10-701-038-1

Query Match   30.2%;
Best Local Similarity 35.7%;
Score 1300.5; DB 16; Length 831;
Pred. No. 1e-83;

```
; PRIOR APPLICATION NUMBER: 09/616,847
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 09/596,193
; PRIOR FILING DATE: 2000-06-17
; PRIOR APPLICATION NUMBER: 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/519,705
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 752
; SEQ ID NO 343
; LENGTH: 1053
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-172-343

Query Match          9.8%;   Score 421;  DB 15;  Length 1053;
Best Local Similarity  25.2%;  Pred. No. 7e-21;
Matches  189;  Conservative  101;  Mismatches  269;  Indels  190;  Gaps  29;

Qy   170 AVLGNKEIKE------------------EWLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIF 215
             ||||              |:||:||||||||: : :||
Db    49 AYLGDAEIIELLILSGARVNAKDSKWLT--PLH-----RAVASCS------EEAVQVLL 94

Qy   216 MNESEIAEGISN---PLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEG 272
             : :::   |||| ||  : | :
Db    95 KHSADVNARDKNWQTPLHIAAANKAVKCAEALVPLLSNV------ 133

Qy   273 DLGANAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKP 332
             ||||         |||||   :  :: ::  :|
Db   134 ------------NVSDRAGRTALHHAAFSGHGEMVKLLLSRGAN--INAFDKKDRRAIHWAAYMG 184

Qy   333 DCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDP 392
             |||         || :|| |    | :::|
Db   185 HIEVVKL--LVSHGAEVTCKDKKSYTPLHAAAS------SGMISVVKYLLDLGVDMNEP 235

Qy   393 DYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIH-VASSNGKANAVSGLV 451
             : :||| || :   :|     | :|
Db   236 NAYGNTPLHVACYN-GQDVVV-NELIDCGAIVNQKNEKGFTPLHFAAASTHGALCLELLV 293

Qy   452 SCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDI 511
             ||||          : ||  :|| |    : :|
Db   294 GNGADVNMKSKDGKTPLHMTALHGRFSRSQTII-QSGAVIDCEDKNGNTPLHIAARYGHE 352

Qy   512 GVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLA 571
             ||||          |: ||||: :: :|
Db   353 LLINTLITS-----GADTAKRGI-----HGMFPLHLAALSGFSD---CCRKLLSSGFDID 399

Qy   572 TI----RAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSN 627
             :|     ||||:| |:||   |::|
Db   400 TPDDFGRTCLHAAAAGGNL------ECLNLLLNTGADF---NKKDKFGRSPLHYAAANCN 450

Qy   628 HNVMKDILN---FVGD-------SLH-----------------LPSERGYNAMQ 654
             | :     |||    ||||:
Db   451 YQCLFALVGSGASVNDLDERGCTPLHYAATSDTDGKCLEYLLRNDANPGIRDKQGYNAVH 510

Qy   655 VAALFGDKEAVKMLAKSAKPSDLNFKTS---------ATPTPLNLACLRGDNEVVRGL 703
             ||:       :::|:   |::
Db   511 YSAAYGHRLCLQLIA-SETPLDVLMETSGTDMLSDSDNRATISPLHLAAYHGHHQALEVL 569

Qy   704 VGQHGIDINQRMGSD-----------------------------KNTVLHYAISKG 730
```

```
US-10-221-278-343
; Sequence 343, Application US/10221278
; Publication No. US20040034208A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc
;  TITLE OF INVENTION: No. US2004003420BA1e1 Nucleic Acids and Polypeptides
;  FILE REFERENCE: 21272-045
;  CURRENT APPLICATION NUMBER: US/10/221,278
;  CURRENT FILING DATE: 2002-09-06
;  PRIOR APPLICATION NUMBER: 09/693,267
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 09/665,363
;  PRIOR FILING DATE: 2000-09-19
;  PRIOR APPLICATION NUMBER: 09/616,847
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: 09/596,193
;  PRIOR FILING DATE: 2000-06-17
;  PRIOR APPLICATION NUMBER: 09/574,454
;  PRIOR FILING DATE: 2000-05-19
;  PRIOR APPLICATION NUMBER: 09/519,705
;  PRIOR FILING DATE: 2000-03-07
;  NUMBER OF SEQ ID NOS: 752
; SEQ ID NO 343
; LENGTH: 1053
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-221-278-343

Query Match          9.8%;   Score 421;  DB 15;  Length 1053;
Best Local Similarity  25.2%;  Pred. No. 7e-21;
Matches  189;  Conservative  101;  Mismatches  269;  Indels  190;  Gaps  29;

Qy   170 AVLGNKEIKE------------------EWLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIF 215
             ||||              |:||:||||||||: : :||
Db    49 AYLGDAEIIELLILSGARVNAKDSKWLT--PLH-----RAVASCS------EEAVQVLL 94

Qy   216 MNESEIAEGISN---PLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEG 272
             : :::   |||| ||  : | :
Db    95 KHSADVNARDKNWQTPLHIAAANKAVKCAEALVPLLSNV------ 133

Qy   273 DLGANAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKP 332
             ||||         |||||   :  :: ::  :|
Db   134 ------------NVSDRAGRTALHHAAFSGHGEMVKLLLSRGAN--INAFDKKDRRAIHWAAYMG 184

Qy   333 DCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDP 392
             |||         || :|| |    | :::|
Db   185 HIEVVKL--LVSHGAEVTCKDKKSYTPLHAAAS------SGMISVVKYLLDLGVDMNEP 235

Qy   393 DYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIH-VASSNGKANAVSGLV 451
             : :||| || :   :|     | :|
Db   236 NAYGNTPLHVACYN-GQDVVV-NELIDCGAIVNQKNEKGFTPLHFAAASTHGALCLELLV 293

Qy   452 SCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDI 511
             ||||          : ||  :|| |    : :|
Db   294 GNGADVNMKSKDGKTPLHMTALHGRFSRSQTII-QSGAVIDCEDKNGNTPLHIAARYGHE 352

Qy   512 GVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLA 571
             ||||          |: ||||: :: :|
Db   353 LLINTLITS-----GADTAKRGI-----HGMFPLHLAALSGFSD---CCRKLLSSGFDID 399

Qy   572 TI----RAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSN 627
             :|     ||||:| |:||   |::|
Db   400 TPDDFGRTCLHAAAAGGNL------ECLNLLLNTGADF---NKKDKFGRSPLHYAAANCN 450

Qy   628 HNVMKDILN---FVGD-------SLH-----------------LPSERGYNAMQ 654
             | :     |||    ||||:
Db   451 YQCLFALVGSGASVNDLDERGCTPLHYAATSDTDGKCLEYLLRNDANPGIRDKQGYNAVH 510

Qy   655 VAALFGDKEAVKMLAKSAKPSDLNFKTS---------ATPTPLNLACLRGDNEVVRGL 703
             ||:       :::|:   |::
Db   511 YSAAYGHRLCLQLIA-SETPLDVLMETSGTDMLSDSDNRATISPLHLAAYHGHHQALEVL 569

Qy   704 VGQHGIDINQRMGSD-----------------------------KNTVLHYAISKG 730
```

RESULT 4

```
Db   49  --AFLGDAEIIELLILSGARVNAKDNMWLT-----PLH----RAVASRS-------EEAVQV  92
         |: |||  |                      :   |||    |||| |          |||||
Qy  214  IFMNESEIAEGISN---PLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIV--KAL-GE  266
         :  : : ||           ||||  : |         ||
Db   93  LIKHSADVNARDKNWQTPLHVAAANKAVKCAEVIIPLLSSVNVSDRGGRTALHHAALNGH  152

Qy  267  VE------------------------NDEGDLGANAY------------NVLDSEGNLPLHH  292
         |:                       |  |: |  :|                 |   :|||
Db  153  VEMVNLLLAKGANINAFDKKDRRALHWAAYMGHLDVVALLINHGAEVTCKDKGKYTPLHA  212

Db  293  AAKNCTGDKLK----------------------LC---MEKTKTDFIDTANFANQ------  322
         || :   |  :                      |    ||  |    ||
Db  213  AASNGQINVVKHLLNLGVEIDEINVYGNTALHIACYNGQDAVVNELIDYGANVNQPNNNG  272

Qy  323  -SPLHIITQKPDCSVLDIEFTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDH---  378
         :|||:                         | || ||| | |  |||||  |  |
Db  273  FTPLHFAAASTH-GAICLELLVNNGADVNIQSKDGKSPLHMTAVH------GRFTRSQT  324

Qy  379  -VKNSSEFQDLVNDPDYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHV  437
          :: :   |  ::   ||||| | ||        ||        ||: ||  |||  |||
Db  325  LIQNGGEIDCVDKD----GNTPLHVAARYGHE--LLINTLITSGADTAKCGIHSMFPLHL  378

Qy  438  ASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNND  497
         |  :   |  ||      ||||   |  |  |||||  |   : | :  |||| | :|||
Db  379  AALNAHSDCCRKLLSSGFEIDTPDKFGRTCLHAAAAGGNVECI-KLLQSSGADFHKKDKC  437

Qy  498  GVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAG------  550
         | || :||| |     |    :  |    || ||     ||  ||  || ||
Db  438  GRTPLHYAAANCHFHCIETIVTTGANVNETD------DW-GRTALHYAAASDMDRNK  487

Qy  551  ---GGSRDTFRVVRKNYEKCHDLATIRAAL------------MQDRSCGGELVN----LG  590
            | ||:|| |   |    | | | ||                :      : :
Db  488  TILGNAHDN---SEBLERARELKEKEATLCLEFLLQNDANPSIRDKEGYNSIHYAAAYG  543

Qy  591  DFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDIINLPVGDSLHILPSERGY  650
            :::| || |    | |    |                               |
Db  544  HRQCLELL---------LERTNSGFEESDSGATKS-------------PLHL------  573

Qy  651  NAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNALCLRGDNEVVRGLVGQHGID  710
            :  ||     |  :|     |                |    |  :   |  |:
Db  574  ------AAYNGHHQALEVLLQSL--VDLDIRDEKGRTALDLAAPKGHTECVEALINQ-GAS  625

Qy  711  INQRMGSDKNTVLHYAISKGDSFLVQKLL--AHTGVDVNCENNLGQTPLHLAVEGGDPKI  768
             : |  || : ||  : ||   | |||    :||||||||| ||||||| |  || |||
Db  626  IFVKDNVTKRTPLHASVINGHTLCLRLLLEIADNPEAVDVKDAKGQTPLMLAVAYGHIDA  685

Qy  769  VSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADI  817
         ||| | || ||: :  |||| | |||::    :  | |:  || :
Db  686  VSLLLEKEANVDTVDILGCTALHRGIMTGHEE----------CVQMLLEQEVSI  729
```

RESULT 6
US-10-108-260A-3237

US-10-108-260A-3237

Query Match   8.5%;
Best Local Similarity 25.1%;   Score 368.5; DB 15; Length 1330;
Matches 178; Conservative 87; Mismatches 262; Indels 183; Gaps 26;

```
Db  570  V-QSLLDLDVRNSSGRTPLDLAAFKGHVECVDVLINQGASILVKDYILKRTPIHAAATNG  628
         | :|:  |  |||||||| ||||||||| ||||||||||:||         |||||||
Qy  731  DSFLVQKILAH---TGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRLDDNG  786
         ::  :                  ||| ||||||: ||| ||||||||||||||:  |
Db  629  HSECLRLLIGNAEPQNAVDIQDGN--GQTPLMLSSVLNGHTDCVYSLLNKGANVDAKDKWG  686

Qy  787  RSVLSSAIVPGRKEKGVLGIVNKLLDRGA  815
         | | ||||                |: ||
Db  687  RTALHRGAVTGHEE-----CVDALLQHGA  710
```

RESULT 5
US-10-479-764-18

US-10-479-764-18

Query Match   8.7%;
Best Local Similarity 23.9%;   Score 373.5; DB 17; Length 1056;
Matches 198; Conservative 101; Mismatches 303; Indels 227; Gaps 35;

```
Qy  108  QPPIVQDSNLYSSICGVPQEAQYDAAARAGGPRKFLYGPYTFSNGQEIMDFEFDTPWPDV  167
         ||| ||         |  |  |                              :
Db    9  QPPLVQ-----AIFSGDPEEI----------RMLIHK-----TEDVNTLDSEKRTPLHVA  48
Qy  168  RNAVLGNKEIKE--------------------EWLTTSGPVRDIADRIVASKGDLSBDQVEEILDI  213
```

```
; OTHER INFORMATION: sequence
US-10-369-978-4

Query Match          8.5%;  Score 368;  DB 14;  Length 1619;
Best Local Similarity  21.9%;  Pred. No. 7.9e-17;
Matches 187;  Conservative 133;  Mismatches 299;  Indels 236;  Gaps 35;

Qy  119  SSIGGVPQEA------QYDAAARAGGPRKFLYGPYTFSNG-QEIMDFEFDTPWP------  165
                     :..:   :                :::..:..: :.  .:
Db   22  SSLTGPPDESATPSERATPASKADSDPK---DDSSSNGDKKDMDL-FPAPKPPSAGASI   76

Qy  166  -DVRNAVLGNKEIKEEWLTTSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNESEIAEG  224
          .  :  :   :  :.:                                .: ::
Db   77  RDTANKVLG-LAMKSEWTPIEAELKKL-EKYVANVGB-------------           111

Qy  225  ISNPLHADVDNNPVKGAKNVMT-LMHLVYAC-DVDPRIVKALGEVENDEGDLGANAYNVL  282
           .:  :       .:.  : .:: :.:.  .   : .: ::       :.  ::
Db  112  ------DGNHIPLAGVHDMNTGMTPLMYATKDNKTAIMDRMELGADVGARNNDNYNV-    163

Qy  283  DSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEE-  341
              .: ::  :..    .  :                       ::.:.:.:.::.::
Db  164  ------LHIAAMYSREDVVKLLLTKRGVDPFSTGGSRSQTAVHLVSSRQTGTATNILRA   216

Qy  342  -FTSRNLDFGL-VDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTI  399
             ..:.: .: .:: : . :.:.::..   .                 .:
Db  217  LLAAAGKDIRLKADGRGKIPLLLAVESGNQSMCRELL------AAQTAEQLKATTANGDTA 271

Qy  400  AHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNS  459
           :     .   .:    :.:.::                .:..: ::      : ::..:
Db  272  LHLAARRR--DVHMVRILVDYGTNVDTQNGEGQTPLHIAAAEGDEALLKYFYGVRASASI  329

Qy  460  QDVNGDTPLHIAVEGGSMETV-----------LAVLN-------Q  486
          .  :.::::                      :.:::       :
Db  330  ADNQDRTPMHLAAENGHAHVIEILADKFKASIFERTKDGSTLMHIASLNGHAECATMLFK  389

Qy  487  RGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIKGEDTV--------AKSLL  535
           :::.::    : ::   ::.:: .    :    :.   ..         :.:.:
Db  390  KGVYLHMPNKDGARSIHTAAAYGHTGIINTLLQKGEKVDVTTNDNYTALHIAVESAKPAV  449

Qy  536  MEDYKGFTPLHFVAGGSRDTFRVVRKNYEKCHDLATI----RAALMQDRSG-------  583
           :::        :.:: : .         :..  .:      ::.::
Db  450  VETLLGFGADVHVRGGKLRET-------------PLHIAARVKDGDRCALMLLKSGASPNLTTD 500

Qy  584  ---------------GELVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHVM  631
                         :.:.: :: :.:       :.  .:         :::
Db  501  DCLTPVHVAARHGNLATIMQLLEDE-GDP---------------LYKSNTGETPLHMACRACHPDIV 551

Qy  632  KDILNFVGDSLHLP---------SERGYNAM----------QVAALFGDKEAVKMLAKSA  672  ●
           .: . :.:.:   :       :.::  :           ::.:::.:..: :: ::
Db  552  RHLIETVKEK-HGPDKATTYINSVNEDGATALHYTCQITKEEVKIPESDKQIVRMLLENG  610

Qy  673  ------------------------KPSD---LNFKTSATPTPLNLACLR  694
                                   ::.:   :..::::::: :: ::. :
Db  611  ADVTLQTKTALETAFHYCAVAGNNDVLMEMISHMNPTDIQKAMNRQSSVGWTPLLIACHR  670

Qy  695  GDNEVVRGLVGQHG-IDINQRMGSDKNTVLHYAISKG-----DSFLVQKILAHTGVDVNC  748
          :..::::.:.:.:   :.:.       .:::.::.       :.::::::::::.::.
Db  671  GHMELVNNLLANHARVDV---FDTEGRSALHLAAERGYLHVCDALLTNKAF------INS  721

Qy  749  ENNLGQTPLHLAVEGGDPKIVSSLLK-AGAVNRLDDNGRSVLSSAIVPGRKEKGVLGIV  807
           ..:: :::::::.::.  .:  ::.  ::::.::.:: : .::                VC 776
Db  722  KSRVGRTALHLAAMNGFTHLVKFLIKDHNAVIDILTLRKQTPLHLAAASGQME-----VC  776

Qy  808  NKLLDRGADINLDGD  822
           ..: ::.::::  :
Db  777  QLILELGANIDATDD  791
```

```
RESULT 8
US-10-108-260A-4122
; Sequence 4122, Application US/10108260A
; Publication No. US2004000550A1
; GENERAL INFORMATION:
```

```
Qy  184  TSGPVRDIADRIVASKGD------LSEDQVEEILDIIFMNESEIAEGISNPLH------  230
          :.:.  :::::::::              ::::::::::::::::
Db  166  TKGKVRLPALHIAARKDDTKSAALLLQNDHNADVQSKMVNRT--TESGFTPLHIAAHYG  223

Qy  231  -ADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKAL----GEVENDEGDL          274
              .:  :: .: .:: :.::.    .  : .:::     :.::::::
Db  224  NVNVATLLLNRGAAVDFT---ARNGITPLH-VASKRGNTNMVKLLLDRGGQIDAKTRD-  277

Qy  275  GANAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDC  334
              :. .::.:: ::: .::. .::                    :.:  :::
Db  278  -+--GLTPLHCAARSGHDQVVELLLERGAPLLARTKN--GLSPLHMAAQ------     320

Qy  335  SVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPV------------ILK         373
             :.  .:.:   :   :: : . ::: .:   :            .:.
Db  .321 -------------GDHVECVKHLLQHKAPVDDVTLDYLTALHVAAHCGHYRVTK       361

Qy  374  GVMDHVQNSSEFQDLVNDPDYFGNTIAHYAVKKNADLTLFNMLKASGADLNVRNVVGRA  433
           :..::.:.      ::.    :.::: : :. :.::::::::.:::::::.:.:. ::
Db  362  LLLLDKRANP-------NARALNGFTPLHIACK--KNR-IKVMELLVKYGASIQAITESGLT  412

Qy  434  PIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSV  493
          :::::. .: .::: ::::::::::: ::.:::::::: :. .:  ::  ::.:::::.:
Db  413  PIHVAAFMGHLNIVLLLLQNGASPDVTNIRGETALHMAAARAGQVEVVRCLL-RNGALVDA  471

Qy  494  QNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNI--------------KGEDTVAKSLL-  535
           :.:::::: :::::::: .  ::  :::::::              ::::::::.::
Db  472  RAREEQTPLHIASRLGKTEIVQLLLQHMAHPDAATTNRYTPLHISAREGQVDVASVLLEA  531

Qy  536  ----MEDYKGFTPLHFVAGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGGELVNLG  590
              ::::::::::::::::::::::::::::                  .:..:  ::
Db  532  GAAHSLATKKGFTPLHVAAKYGSLDVAKL--------------LLQRAAAD-         569

Qy  591  DFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGY  650
          .:::::::::::::::::::::::::::::::::::.:::               :.::.
Db  570  --------SAGKNGLTPLHVAAHYDNDQKVALLLLE-KGASPHATAKNGY          609

Qy  651  NAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGID  710
           :..:::::::::::::::::                :::::::::::::::::::.:::
Db  610  TPLHIAAKKNQMQIASTLLNYG--AETNIVTKQGVTPLHLASQEGHTDMVTLLLDK-GAN  666

Qy  711  INQRMGSDKN--TVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKI  768
           ...  :.    .::: ::.:::::::::::::::::::.:::::::::::::::::::::
Db  667  IHM--STKSGLTSLHLA-AQEDKVNVADIILTKHGADQDAHTKLGYTPLIVACHYGNVKM  722

Qy  769  VSSLLKAGAVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADIN          818
          : .:::.:.::: .::  ::::.::.    :.   .. .:::.:::::
Db  723  VNFLLKQGANVNAKTKNGYTPLHQA-----AQQGHTHIINVLLQHGAKPN         767
```

```
RESULT 7
US-10-369-978-4
; Sequence 4, Application US/10369978
; Publication No. US2003015291A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Walker, Richard G.
; APPLICANT: Willingham, Aarron
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: A Eukaryotic Mechanosensory Transduction Channel
; FILE REFERENCE: 02307E-097600US
; CURRENT APPLICATION NUMBER: US/10/369,978
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US/09/392,812A
; PRIOR FILING DATE: 1999-09-09
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
; LENGTH: 1619
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: amino acid sequence derived from nompC cDNA
```

```
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; SOFTWARE: PatentIn Ver. 2.1
; NUMBER OF SEQ ID NOS: 5458
; SEQ ID NO 4122
; LENGTH: 919
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4122

Query Match      8.5%;    Score 364.5;   DB 15;   Length 919;
Best Local Similarity    24.0%;   Pred. No. 6.1e-17;
Matches  184;  Conservative  95;  Mismatches  283;  Indels  205;  Gaps  31;

Qy 170 AVLGNKEIKE----------------EWLTTSGPVRDIADRIVASKGDLSEDQVEILDIIF 215
                             :  ::
Db  24 AFLGDAEIIELLILSGARVNAKDNMWLT---PLH----RAVASRS------EEAVQVLi   69

Qy 216 MNESEIAEGISN---PLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIV--KAL-GEVE  268
       : :::            ||||                                       
Db  70 KHSADVNARDKNWQTPLHVAAANKAVKCAEVIIPLLSSYNVSDRGGRTALHHAALNGHVE 129

Qy 269 --------------------------NDEGDLGANAY----------NVLDSEGNLPLHHAA 294
                                                          ||  |||||||
Db 130 MVNLLLAKGANINAFPDKKDRRALHWAAYMGHLDVVALLINHGAEVTCKDKKGYTPLHAAA 189

Qy 295 KNCTGDKLK------------------------LC----MEKTKTDFIDTANFANQ------S 323
       |||                                                         
Db 190 SNGQINVVKHLLNLJGVEIDEINVYGNTALHIACYNGQDAVVNELIDYGANVNQPNNNGFT 249

Qy 324 PLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDH----V 379
       |||                           |||||                         
Db 250 PLHFAAASTH-GAlCLELLVNNGADVNIQSKDGKSPLHMTAVH----------GRFTRSQTLi 301

Qy 380 KNSSEFQDLVNDPDYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVAS 439
                   ||||                   |||||           :   || ||
Db 302 QNGGGEIDCVDKD------GNTPLHVAARYGHE--LLINTLRTSGADTAKCGIHSMFPLHAA 355

Qy 440 SNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGV 499
                                   ||||                            
Db 356 LNAHSDCCRKLLSSGFEIDTPDKFGRTCLHAAAAGGNVECI-KLLQSSGADFHKKDKCGR 414

Qy 500 TPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAG---------550
       || :                                          ||||||        
Db 415 TPLHYAAANCHFHCIETLVTTGANVNETD----------DW-GRTALHYAAASDMDRNKTi 464

Qy 551 -GGSRDTFRVVRKNYEKCHDLATIRAAL----------MQDRSGGELVN----LGDF   592
         ::  :                                  |||||||          
Db 465 LGNAHDN----SEELERARELKEKEATLCLEFLLQNDANPSIRDKEGYNSIHYAAAYGHR 520

Qy 593 ESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNPVGDSLHLPSERGYNA 652
       :                                                     |||    
Db 521 QCLELL----------LERTNSGFPEESDSGATKS--------------------PLHL---548

Qy 653 MQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGIDIN 712
        ::  |                                                      
Db 549 ---AAYNGHHQALEVLLQSL--VDLDIRDEKGRTALDLAAFKGHTECVEALINP-GASIF 602

Qy 713 QRMGSDKNTVLHYAISKGDSFLVQKIL--AHTGVDVNCENNLGQTPLHLAVEGGDPKIVS 770
           ||   ||| :                        ||         ||         
Db 603 VKDNVTKRTPLHASVINGHTLCLRLLLEIADNPEAVDVDAKGQTPLMLAVAYGHIDAVS 662

Qy 771 SLLKAGAVVNRLDDNGRSVLLSSAIVPGRKEKGVLGIVNKLLDRGADI 817
                                                ||  |||   
Db 663 LLLEKEANVDTVDILGCTALHRGIMTGHEE------CVQMLLEQEVSI 704


RESULT 9
US-10-369-978-6
; Sequence 6, Application US/10369978
```

```
; Publication No. US20030152991A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Walker, Richard G.
; APPLICANT: Willingham, Aarron
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: A Eukaryotic Mechanosensory Transduction Channel
; FILE REFERENCE: 02307E-09760US
; CURRENT APPLICATION NUMBER: US/10/369,978
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US/09/392,812A
; PRIOR FILING DATE: 1999-09-09
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 6
; LENGTH: 1709
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
; OTHER INFORMATION: amino acid sequence derived from nompC genomic
; OTHER INFORMATION: sequence
US-10-369-978-6

Query Match      8.1%;    Score 351;   DB 14;   Length 1709;
Best Local Similarity    21.1%;   Pred. No. 1.4e-15;
Matches  196;  Conservative  134;  Mismatches  326;  Indels  272;  Gaps  38;

Qy 108 QPPIVQDSNLYSSIGG----------VPQEAQYDAAA--RAGGPRKFLYGPYTFSNGQEIMDF 158
       :||:  ::: :                        ::::                       
Db  50 RPPL-RRSSTHLQIGKNSRIIFVKQPSRDSVTPPDRLLGKPLFRESLTSHASSHEEMSS 108

Qy 159 EFDTPWPDVRNAVLGNKEIKEEWLTTSGPVRDI----------ADR----IVASKGDL 202
                                                                    
Db 109 E-DLAMADPQTKILYFAK-RDEWANVESEIETIKRSDFSMADNHGFTAFLLAVKAGKDQI 166

Qy 203 SEDQVE----------------------EILDIIFMNESEIAEGISNP----LHAD 232
         ::                                                        
Db 167 VDKMIRKGARVDYSTKDGRNATHIAAMYSGVETLELILKRYSELLRKGAGPKKQLAIHVA 226

Qy 233 VDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDL----GANAYNVL-   282
                              :                                    
Db 227 CERKSKKAFPIVKRILE----DTDQRM--------AEDGDGSLPIHLAFKFGNVNIVELLL 275

Qy 283 --------------DSEGNLPLHHAAKNCTGDKLKLCMEKTKTDF---IDTANFAN--QSPL 325
                                                                  :: 
Db 276 SGPSDEQTRKADGNGDTLLHLAAR----SGN----IEAVRTAIAAGCDNANVQNRVGRTPL 328

Qy 326 HIITQKPDCSVLD----------IEEFTSRNL------DFGLVDGDGKNPLHHAVEHLPPV 370
                                                                    
Db 329 H------ECLTVTGTQKGVVAEVGDQNMLKIMFKLRADANIHDKEDKTPVHVAAERGDTS 382

Qy 371 ILKGVMDHVKNSSEEFQDLVNDPDYFGNTIAHYA----------VKNKNADL 411
                                   ::::                           
Db 383 MVESLID------KFGGSIRARTRDGSTLLHIAACSGHTSTALAFLKRVPLFMPNKKGAL 436

Qy 412 TL-----FN----MLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVN--458
       |                                                          
Db 437 GLHSAAAAGFNDVVKMLIARGTNVDVRTRDNYTALHVAVSGKASVVETLLGSGADIHVK 496

Qy 459 -SQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL 517
                     :                                              
Db 497 GGELMDGETCLHIAARSGNKDIML--LLDENADSKISSKIGETPLQVAAKSCNFEAASMI 554

Qy 518 GSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGGGSR------DTFRVVRK--NYEKCHD 569
                                 |                                   
Db 555 LKHILSEVLTQEQLKEHVNHRTNDGFTALHYAAEIEQRQLHFPGBDAKLVNLLIDYGGMVE 614

Qy 570 LATIR----AALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVS 625
            :                                                       
Db 615 MPSLNANETAMHMAARSGNQAVLLA--MVNKIGAGAVQIVQNKQSKN-GWSPLLEACAR 670

Qy 626 SNHNVMKDIL---- 635
         :      :|   
Db
```

```
Db   671 GHSGVANILLKVLVLCVGPGPGPGRLQGRGYWTRTRARVTVPWLQYQGYWARTRTRA 730

Qy   636 ------------------------NFVGD---SLHLPSERGYNAMQVAALFGDKEAVKM 667
                                 :||| :|:||||||||||:||||  :|||:|||
Db   731 RATGPGLQDQGYWARTRTRKVTVPRLLGDHHARIDVFDEMGRTALHLAAFNGHILSLVHL 790

Qy   668 LAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDKNTVLHYAI 727
             ||||| |:|||||||||||||||||||||||||||| :|||||||||||| ||||||
Db   791 LLQ--HKAFVNSKSKTGEAPLHLAAQHGHVKVVNVLVQDHGAAL-EAITLDNQTALHFAA 847

Qy   728 SKGDSFIVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKA----GAVNRLD 783
             ||||| ||: | ::||| :|||||||||||| |||
Db   848 KFGQLAVSQTLLA-LGANPNARDDKGQTPLHLAAENDFPDVVKLFLKMRNNNRSVLTAID 906

Qy   784 DNGRSVLSSAIVPGRKEKGVLGIVNKLL 811
             : :||||||||||: :|:|
Db   907 HNGFTCAHIAAM-----KGSLAVVRELM 929
```

RESULT 10
US-10-479-764-22
; Sequence 22, Application US/10479764
; Publication No. US2005002710A1
; GENERAL INFORMATION:
;   APPLICANT: TANG, Y. Tom; WARREN, Bridget A.;
;   APPLICANT: HONCHELL, Cynthia D.; RICHARDSON, Thomas W.;
;   APPLICANT: ELLIOTT, Vicki S.; CHAWLA, Narinder K.;
;   APPLICANT: YUE, Henry; BATRA, Sajeev;
;   APPLICANT: GRIFFIN, Jennifer A.; BAUGHN, Mariah R.;
;   APPLICANT: FORSYTHE, Ian J.; BURFORD, Neil;
;   APPLICANT: EMERLING, Brooke M.; SANJANWALA, Madhusudan M.;
;   APPLICANT: KHAN, Farrah A.; LU, Dyung Aina M.;
;   APPLICANT: HAFALIA, April J.A.; NGUYEN, Danniel B.;
;   APPLICANT: YANG, Junming; LI, Joana X.;
;   APPLICANT: BECHA, Shanya D.; YAO, Monique G.;
;   APPLICANT: GIETZEN, Kimberly J.; LUO, Wen;
;   APPLICANT: LEE, Ernestine A.; ISON, Craig H.;
;   APPLICANT: LASEK, Amy W.
;   TITLE OF INVENTION: STRUCTURAL AND CYTOSKELETON-ASSOCIATED PROTEINS
;   FILE REFERENCE: PF-1007 USN
;   CURRENT APPLICATION NUMBER: US/10/479,764
;   CURRENT FILING DATE: 2003-12-04
;   PRIOR APPLICATION NUMBER: PCT/US02/17956
;   PRIOR FILING DATE: 2002-06-06
;   PRIOR APPLICATION NUMBER: US 60/296,865
;   PRIOR FILING DATE: 2001-06-07
;   PRIOR APPLICATION NUMBER: US 60/296,878
;   PRIOR FILING DATE: 2001-06-08
;   PRIOR APPLICATION NUMBER: US 60/298,664
;   PRIOR FILING DATE: 2001-06-15
;   PRIOR APPLICATION NUMBER: US 60/300,149
;   PRIOR FILING DATE: 2001-06-21
;   PRIOR APPLICATION NUMBER: US 60/302,340
;   PRIOR FILING DATE: 2001-06-29
;   PRIOR APPLICATION NUMBER: US 60/303,481
;   PRIOR FILING DATE: 2001-07-06
;   PRIOR APPLICATION NUMBER: US 60/305,059
;   PRIOR FILING DATE: 2001-07-12
;   PRIOR APPLICATION NUMBER: US 60/343,557
;   PRIOR FILING DATE: 2001-12-01
;   NUMBER OF SEQ ID NOS: 50
;   SOFTWARE: PERL Program
; SEQ ID NO 22
;   LENGTH: 1094
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID NO: 8013295CD1
US-10-479-764-22

Query Match   8.1%;   Score 350;   DB 17;   Length 1094;
Best Local Similarity   25.7%;
     Pred. No. 8.5e-16;

```
Matches 127;  Conservative  75;  Mismatches 206;  Indels 86;  Gaps 17;

Qy   323 SPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNS 382
             :||||  |:||                          :|:||
Db    23 TPLHIACKKNRIKVMEL--LLKHGASIQAVTESGLTPIHVA--------AFMGHVNIV 70

Qy   383 SEFQD---LVNDPDYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVAS 439
             :| :         |:|||||||||||:||:|||||| ::|||||:||||||||||||
Db    71 SQLMHGASPNTNVRGETALHMAARSGQAEVVRY--LVQDGAQVEAKADQTPLHISA 128

Qy   440 SNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGV 499
             |:|||:|||| |||||||:|||:|||||||||||| |||||||:|||:|||||||||||
Db   129 RLGKADIVQQLLQQGASPNAATTSGYTPLHLSAREGH-EDVAAFLLDHGASLSITTKKGF 187

Qy   500 TPMLSAAKYGDIGVIKAL--GSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTF 557
             ||| |||:|||||||:|   |:| ||| ||||||||||:||||||||||:|   :||||
Db   188 TPLHVAAKYGKLEVANILLQKSASPDAAGK-----SGLTPLHVAA---HYDNQ 232

Qy   558 RVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGSPNAKPLQHIQSANFGFS 617
             :              :     |||||||
Db   233 KV----------ALLLDQG--------ASPHA--------AAKNGYT 254

Qy   618 PARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAKSAKPSDL 677
             ||||||||||||:|||| |||||||||||||||||||||||||||||||||||||:||
Db   255 PLHIAAKKNQMDIATTLLEYGADA-NAVTRQGIASVHLAAQEGHVDMVSLLL--GRNANV 311

Qy   678 NFKTSATPTPLNLACLRGDNEVVRGLVGQHG-IDINQRMGSDKNTVLHYAISKGDSFLVQ 736
             |||||||||||||||||| | |||||:|||| |||||||||||:|||||||||||:|:|
Db   312 NLSNKSGLTPLHLAAQEDRVNVAEVLVNQGAHVDAQTKMG---YTPLHVGCHYGNIKIVN 368

Qy   737 KILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVNRLDDNGRSVLSSAIVP 796
             :||||:||||||||||:|||||||:|||||||||||:|||:| ||: |:|| ||||||
Db   369 FLLQHS-AKVNAKTKNGYTPLHQAAQQGHTHIINVLLQNNASPNELTVNGNTALGIA--- 424

Qy   797 GRKEKGVLGIVNKLL 810
             |::|| |||||:|||
Db   425 --RRLGYISVVDTL 436
```

RESULT 11
US-10-205-194-117
; Sequence 117, Application US/10205194
; Publication No. US2003013430IA1
; GENERAL INFORMATION:
;   APPLICANT: Warner-Lambert Company
;   APPLICANT: Lee, Kevin
;   APPLICANT: Dixon, Alistair
;   APPLICANT: Brooksbank, Robert
;   APPLICANT: Pinnock, Robert
;   TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
;   FILE REFERENCE: WL-A-018201
;   CURRENT APPLICATION NUMBER: US/10/205,194
;   CURRENT FILING DATE: 5200-07-24
;   PRIOR APPLICATION NUMBER: GB 0118354.0
;   PRIOR FILING DATE: 2001-07-27
;   NUMBER OF SEQ ID NOS: 177
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 117
;   LENGTH: 1762
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Ankyrin isoform
US-10-205-194-117

Query Match   8.0%;   Score 345.5;   DB 14;   Length 1762;
Best Local Similarity   24.8%;   Pred. No. 3.6e-15;
Matches 169; Conservative 95; Mismatches 284; Indels 133; Gaps 26;

```
Qy   184 TSGPVRDIADRIVASKGDLSEDQVEBILDIIFMNES----EIAEGISNPLH----------- 230
             |   :|::|| :||:||              :|: |    : :::|||
Db   191 TKGKVRLPALHIAARKDD------TKAAALLQNDTNADIESKSGFT-PLHIAAHYGNIN 243
```

```
;   PRIOR FILING DATE: 2000-12-14
;   PRIOR APPLICATION NUMBER: US 60/256,172
;   PRIOR FILING DATE: 2000-12-15
;   PRIOR APPLICATION NUMBER: US 60/257,416
;   PRIOR FILING DATE: 2000-12-22
;   PRIOR APPLICATION NUMBER: US 60/260,912
;   PRIOR FILING DATE: 2001-01-10
;   PRIOR APPLICATION NUMBER: US 60/264,644
;   PRIOR FILING DATE: 2001-01-25
;   PRIOR APPLICATION NUMBER: US 60/266,017
;   PRIOR FILING DATE: 2001-02-02
;   NUMBER OF SEQ ID NOS: 40
;   SOFTWARE: PERL Program
;   SEQ ID NO 20
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   OTHER INFORMATION: Incyte ID No: 3566882CD1
US-10-433-794-20

Query Match             8.0%;  Score 344;  DB 15;  Length 720;
Best Local Similarity  21.9%;  Pred. No. 1.2e-15;
Matches  178;  Conservative 113;  Mismatches 269;  Indels 254;  Gaps  33;

Qy   80  PTLPP-------RTFSGE------------GYDDVGVS--MPTVSRGIYQPPIVQDSNLYSSIGGV  124
                       :  :         :::: ::: ||      ||:
Db   54  PCLPPDAARTFAASVSPLPSIYLAKISDFGLSKWMEQSTRMQY---IERSALRGMLSYI        109

Qy  125  PQEAQYDAAARAGGPRKFLYGPYTF-SNGQEIMDFEFDTPWPDVRNAVLGNKE------------  176
              :                ::  |:||        : ::|: :
Db  110  PPE-MFLESNKAPGPKYDVYSPPTLPPRAGVILDVQ----------------LSHSERVLCIHS  156

Qy  177  ---IKEEWLTTSGVRDIADRIVASKGDLSEDQVEBILDIIFMNESEIAEGISNPLHADV        233
            ::|        ::                                  :
Db  157  FAIVLWELLTQKKPYSELTSQLKERKG------------------FNMMMII----RVTAGMRPSLQPVS  205

Qy  234  DNNPVKGAKNVMTLMHLVTACDVDP-----------------------RIV----KAL--        264
            |||                                             ::    :
Db  206  DQWFSE-AQQMVDLMKRCW---DQDPKKRPCFLDITIETDILLSLLQSRVAVPESKALARK  262

Qy  265  ------GEVENDEGDLGANAYNVLDSEGNLPLHHA-----AKNCTGDKLKLCMEKT        309
                  ::
Db  263  VSCKLSLRQPGEVNED---------ISQELMDSDSGNYLKRALQLSDRKNLVPRDEELJCIYEN  316

Qy  310  KTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPP  369
         |||                         ::|| |
Db  317  KV----------TPLHFLVAQG--SVEQVRLLLAHEVDVDCQTASGYTPLLIAAQDDQQP  363

Qy  370  VILKGVMDHVKNSSEFQDLVNDPDYFGNTIAHYAVKNKNADLTLFMLKASGADLNVRNV  429
                                   |||:
Db  364  -------------------------------DLCALLAHGADANRVDE           381

Qy  430  VGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGA  489
              | |||         ::: || :  || ||||||           :   |
Db  382  DGWAPLHFAAQNGDDGTARLLLDHGACVDAQEREGWTPLHLAAQ-NNFENVARLLVSRQA  440

Qy  490  DVSVQNNDGVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVA  549
          :|| | ||||| :||||:  |  :| :    |   |||:    :|::||: |||||| :
Db  441  DPNLHEAEGKTPLHVAAYFGHVSLVKLLTS----QGAELDAQQRNLR----TPLHLAV   490

Qy  550  GGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGSPNAKFLQHI  609
         |                            :| ||
Db  491  ERGK------------VRA-----------------------------IQHL         501

Qy  610  QSA------NFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGD  661
                  :      |::       |:|::|     ::|   |:
Db  502  LKSGAVPDALDQSGYGPLHTAAAARGKYLICKMLLRY-GASLELPTHQGWTPLHLAAYKGH  560

Qy  662  KEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDKN-  720
         :||:                  |           ||  :  | :
Db  561  LEIIHLLAES--HANMGALGAVNWTPLHLAARHGEAVVSALL------QCGADPNA     609
```

Left column:

```
Qy  231  ----------ADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNV  281
          ||                      ||    :   :||
Db  244  VATLLLNRAAAVDFT----ARNDITPLH-VASKRGNANMVKLL--------LDRGAK-IDA  290

Qy  282  LDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEE  341
          ||:                         ||:::                      ::
Db  291  KTRDGLTPLHCGARSGHEQVVEMLLDRAAPILSKTKN--GLSPLHMATQGDHLNCVQL--  346

Qy  342  FTSRNLDFGLVDGDGKNPLHHAVEHLPPVVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAH  401
         |||                  |  :                            ||
Db  347  LLQHNVPVDDVTNDYLTALHVAAHCGHYKVAKVLLDKKANP------NAKALNGFTPLH  399

Qy  402  YAVKNKNADLTLFNMLKASGADLNRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQD  461
          || |                ||   ||||||              |:
Db  400  IACK-KNR-IRVMELLLKHGASIQAVTESGLTPIHVAARFMGHVNIVSQLMHHGASPNTTN  457

Qy  462  VNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GS  519
         |||||||| :    :    :|  | |||
Db  458  VRGETALHMAARSGQAE-VURYLVQDGAQVEAKAKDEQHPTPHFSRLGKAEIVQQVLQQG  516

Qy  520  AKPNI-------KGEDTVAKSIL------MEDYKGFTPLHFVAGGGSRDTFR  558
         |:|:                 |      |::|||||||||
Db  517  ASPNAATTSGYTPLHLSAREGHEDVAAFLLDHGAFLSITTKKGFTPLHVAAKYGKLEV--  574

Qy  559  VVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSP  618
         ||                    :||
Db  575  ------------ASLLLQKS-------------------ASPDA------AGKSGLTP  595

Qy  619  ARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAKSAKPSDLN  678
         || |                ||    :   |:     ||:
Db  596  LHVAAHYDNQKVALLLLD-QGASPHAAAKNGYTPLHIAAKKNQMDIATSLLEYG--ADAN  652

Qy  679  FKTSATPTPLNLACLRGDNEVVRGLVGQH-GIDINQRMGSDKNTVLHYAISKGDSFLVQK  737
         |||                :  |   ||  :||
Db  653  AVTRQGIASVHLAAQEGHVDMVSLLSRNANVLSNKRGLNP--LHLG-GQEDRVNVAE  708

Qy  738  ILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPG  797
         | |                 ||                            :|
Db  709  VLVNQGAHVDAQTKMGYTPLHVGCHYGNIKIVNFLLQHSAKVNAKTKNGYTPLHQA----  764

Qy  798  RKEKGVLGIVNKLLDRGADIN  818
          |:                ||
Db  765  -AQQGHTHINVLLQNNASPN  784
```

```
RESULT 12
US-10-433-794-20
;   Sequence 20, Application US/10433794
;   Publication No. US2004007044A1
;   GENERAL INFORMATION:
;   APPLICANT: YUE, Henry; DING, Li;
;   APPLICANT: LAL, Preeti G.; GRIFFIN, Jennifer A.;
;   APPLICANT: GURURAJAN, Rajagopal; BAUGHN, Mariah R.;
;   APPLICANT: ISON, Craig H.; RAMKUMAR, Jayalaxmi;
;   APPLICANT: TRIBOULEY, Catherine M.; SWARNAKAR, Anita;
;   APPLICANT: BURFORD, Neil; BANDMAN, Olga;
;   APPLICANT: THORNTON, Michael; KHAN, Farrah A.;
;   APPLICANT: WALIA, Narinder K.; NGUYEN, Danniel B.;
;   APPLICANT: ELLIOTT, Vicki S.; XU, Yuming;
;   APPLICANT: YAO, Monique G.; GANDHI, Ameena R.;
;   APPLICANT: ARIZU, Chandra S.; FORSYTHE, Ian J.
;   TITLE OF INVENTION: KINASES AND PHOSPHATASES
;   FILE REFERNCE: PI-0311 USN
;   CURRENT APPLICATION NUMBER: US/10/433,794
;   CURRENT FILING DATE: 2003-06-04
;   PRIOR APPLICATION NUMBER: PCT/US01/47431
;   PRIOR FILING DATE: 2001-12-04
;   PRIOR APPLICATION NUMBER: US 60/254,034
;   PRIOR FILING DATE: 2000-12-06
;   PRIOR APPLICATION NUMBER: US 60/251,814
;   PRIOR FILING DATE: 2000-12-07
;   PRIOR APPLICATION NUMBER: US 60/255,756
```

```
Qy   714  RMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLL  773
             |||||| :| :: ||||||||||||||
Db   457  ATTSG-YTPLHLSAREGHEDVAAFLLDH-GASLSITTKSGLTPLHVAAHYDNQKVALLLL  514

Qy   774  KAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADIN  818
             | ||||
Db   515  DQGASPHAAAKNGYTPLHIA-----AKKNQMDIATTLLEYGADAN  554
```

RESULT 14
US-10-182-243-56
; Sequence 56, Application US/10182243
; Publication No. US20040048310A1
; GENERAL INFORMATION:
;   APPLICANT: PLOWMAN, GREGORY D.
;   APPLICANT: WHYTE, DAVID
;   APPLICANT: MANNING, GERARD
;   APPLICANT: SUDARSANAM, SUCHA
;   APPLICANT: MARTINEZ, RICARDO
;   TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES AND PROTEIN KINASE-LIKE
;   TITLE OF INVENTION: ENZYMES
;   FILE REFERENCE: 038602/1366
;   CURRENT APPLICATION NUMBER: US/10/182,243
;   CURRENT FILING DATE: 2003-07-07
;   PRIOR APPLICATION NUMBER: PCT/US01/02337
;   PRIOR FILING DATE: 2001-01-25
;   NUMBER OF SEQ ID NOS: 84
;   SOFTWARE: PatentIn Ver. 2.1
;
;   SEQ ID NO 56
;   LENGTH: 765
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-182-243-56

  Query Match   7.6%;   Score 327; DB 15;   Length 765;
  Best Local Similarity   21.4%;   Pred. No. 2.2e-14;
  Matches 161; Conservative 121; Mismatches 262; Indels 208; Gaps 28;

```
Qy    94  DVGVS--MPTVSRGIYQPPIVQDSNLYSSIGGVPQEAYDAAARAGGPRKFLYGPYTFSN  151
             | || | : || ::| :| : ::  ||:|
Db   163  DFGLSKWMEQSTRMQY----IERSALRGMLSYIPPE-MFLESNKAPCGPK---YDVYSFA-  213

Qy   152  GQEIMDEFDT----PWPDVRNAVL------------GNKEIKEEWLTTSGPVRDIADRIVAS  198
             :|| |:     :|  |  |  ||
Db   214  ---IVIWELLITQKKPYSGFNMMMIIIRVAAGMRPSLQPVSDQWPSEAQQMVDLMKRC---  267

Qy   199  KGDLSEDQVEEILDI------------IFMNESEIAEGISNPLHADVDNNPVKGAKNVMTLMHLV  251
             :| | | |||  |  :  : | |||
Db   268  -WDQDPKKRPCFLDITITETDILLSLLQSRVAVPESKAL----------ARKVSCKLSLR  315

Qy   252  YACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKT  311
             ||: |  : |||| :: ||
Db   316  QPGEVENEDISQEL--MDSDSGNYLKRALQLSDRKNLVPRDE----------ELCIYENKV  363

Qy   312  DFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPVI  371
Db   364  ------------TPLHFLVAQG--SVEQVRLLLAHEVDVDCQTASGYTPLLIAAQDQQP--  408

Qy   372  LKGVMDHVKNSSEFQDLVNDPDYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVG  431
Db   409  ----------------------------------------------DLCALLLAHGADANRVDEDG  428

Qy   432  RAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADV  491
             :||| :  || :|| |:|| |  ||::
Db   429  WAPLHFAAQNGDDGTARLLLDHGACVDAQEREGWTPLHLAAQ-NNFENVARLLVSRQADP  487

Qy   492  SVQNNDGVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKSLLMEDYKGFTPLHFVAGG  551
             : |  |  ||| | ||:| |: :| | |     |:||
Db   488  NLHEAEGKTPLHVAAYFGHVSLVKLLTS-----QGAELDAQQRNLR---TPLHLAVER  537

Qy   552  GSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGSPNAKPLQHIQS  611
             ||
Db   538  GK-------------VRA-------------------------------------IQHLLK  548
```

RESULT 13
US-09-964-899-43
; Sequence 43, Application US/09964899
; Patent No. US20020174446A1
; GENERAL INFORMATION:
;   APPLICANT: Cohen, Dalia et al.
;   TITLE OF INVENTION: Identification of Genes Involved in
;   TITLE OF INVENTION: Alzheimer's Disease Using Drosophila Melanogaster
;   FILE REFERENCE: 4-31612 A
;   CURRENT APPLICATION NUMBER: US/09/964,899
;   CURRENT FILING DATE: 2001-09-27
;   PRIOR APPLICATION NUMBER: 60/236,893
;   PRIOR FILING DATE: 2000-09-29
;   PRIOR APPLICATION NUMBER: 60/298,309
;   PRIOR FILING DATE: 2001-06-14
;   NUMBER OF SEQ ID NOS: 53
;   SOFTWARE: FastSEQ for Windows Version 4.0
;
;   SEQ ID NO 43
;   LENGTH: 1724
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-964-899-43

  Query Match   8.0%;   Score 343; DB 9;   Length 1724;
  Best Local Similarity   24.3%;   Pred. No. 5.3e-15;
  Matches 157; Conservative 101; Mismatches 263; Indels 124; Gaps 27;

```
Qy   197  ASKGDLSE--DQVEEILDIIFMNESEIAEGISNPLH----------------ADVDNN  236
Db    11  ARAGHLEKALDYIKNGVDINICNQNGLU-----NALHLASKEGHVEVSELLQREANVDAA  65

Qy   237  PVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKN  296
Db    66  TKKG---NTALHIASLAG-QAEVVKVL-VTN------GANV-NAQSQNGFTPLYMAAQE  112

Qy   297  CTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDG  356
Db   113  NHLEVVKFLLDNGASQSLATED--GFTPLAVALQQGHDQVVSL------ILLENDT  159

Qy   357  KNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDY---FGNTIAHYAVKNKNADLTL  413
Db   160  KGKV-----RLPALHIAARKDDTKAAALLLQNDNNADVESKSGFTPLHIAAHYGNIVAT  214

Qy   414  FNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVE  473
Db   215  LLLNRAAAVDFTARNDI--TPLHVASKRGNANMVKLLLDRGAKIDAKTRDGLTPLHCGAR  272

Qy   474  GGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKALGSAKPNIKGEDTVAKS  533
Db   273  SGH-EQVVEMLLDRAAPILSKTKNGLSPLHMATQGDHLNCVQLL--LQHNVPVDD----  324

Qy   534  LLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFE  593
Db   325  -VTNDY--LTALHVAAHCGH---YKV---------------AKVLLDKK-----  352

Qy   594  SENILGSPNAKPLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAM  653
Db   353  ----ANPNAKALN------GFTPLHIACKNRIKVMELLLKH-GASIQAVTERGETAL  399

Qy   654  QVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGIDINQ  713
Db   400  HMAARSGQAEVVRYLVQDG--AQVEAKADDQTPLHISARLGKADIVQQLL-QQGASPNA  456
```

```
Qy   721  ------TVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLK  774
                  ||:| :| :  : ::  |  ||||||||||||||
Db   610  AEQSGWTPLHLAVQRSTFLSVINLLEH-HANVHARNKVGWTPAHLAALKGNTAILKVLVE  668

Qy   775  AGAVVNRLD-DNGRSVLSSAIVPGRKEKGVLGIV  807
             |||      :| ||      :::|| |::|
Db   669  AGA---QLDVQDGVSCTPLQLALRSRKQGIMSFL  699
```

Sun Feb 27 11:16:08 2005 us-10-701-038-2.rapb Page 10

```
Qy   612 A-------NFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKE 663
              :|| |:|||                         ::  :  ||| :||  ::|
Db   549 SGAVPDALDQSGYGPLHTAAARGKYLICKMLLRY-GASLELPTHQGWTPLHLAAYKGHLE 607

Qy   664 AVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDKN--- 720
              ||:|::|                      ||| |||||||||||||||||||:|:
Db   608 IIHLLAES--HANMGALGAVNWTPLHLAARHGEEAVVSALL---------QCGADPNAAB 656

Qy   721 ---TVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAG 776
                 |||  |||:|:    ::||||    :|||||||||||||||:|||||    :||
Db   657 QSGWTPLHLAVQRSTFLSVINLLEH-HANVHARNKVGWTPAHLAALKGNTAILKVLVEAG 715

Qy   777 AVVNRLD-DNGRSVLSSAIVPGRKEKGVLGIV 807
              : :::       : :::||::: :
Db   716 A--QLDVDGVSCTPLQLALRSRKQGIMSFL 744
```

```
Qy   181 WLTTSG-PVRDIADRIVASKGDLSEDQVEILDIIFMNESEIAEGISNPLHADVDNN---- 236
              :|| |||:|| || :|      :  : :: ||   |    : :|  |
Db   295 WMIWNGTPVRDSLSTLIPKS-----------------QEVLQLL----------VKAGAHVNSEDDRTS 336

Qy   237 -----PVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLH 291
                   :   :||        :|: |:    || |                  :|  |:
Db   337 CIVRQALEREDSIRTLL-------------------------------DNGASV-NQCDSNGRTLLA 371

Qy   292 HAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGL 351
              :|:| |||:| ||   :  ::|| |   :||: || :: :     :|| :::||
Db   372 NAAYSGSLDVNLLVSRGADLEIEDAH--GHTPLTLAARQGHTKVVNCLIGCGANINH-- 427

Qy   352 VDGDGKNPLHHA----VEHLPPVILKGVM- 385
              |  ||  |||                    :
Db   428 TDQDGWTALRSAAWGGHTEVVSALLYAGVKVDCADADSRTALRAAAWGGHEDIVLNLLQH 487

Qy   386 QDLVNDPDYFGNTIAHYAVQNKNADLTLFNMLKASGADLNVRNVVGRAPIHVA------SS 440
              |   |   |||:|||||:||:| |||   : |: |||  |::  ||: | :|     ::
Db   488 GAEVNKADNEGRTALIAAAYMGHREIV-EHLLDHGAEVNHEDVDGRTALSVAALCVPAS 545

Qy   441 NGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVT 500
              :|  | |:||||| ||:  |:|   ||||:||   : ||: |::  |||  ::::|  :
Db   546 KGHASVVSLLIDRGAEVDHCDKDGMTPLLVAAYRGHVD-VVDLLEEGADVDHTDNNGRT 604

Qy   501 PMLSAAKYGDIGVIKAL-----GSAKPNI--------------KGEDTVAKSLL-------ME 537
              |  ||| |||   |||      ||  ||               |||| |:||
Db   605 PLLAAASMGHASVVNTLLFWGAAVDSIDSEGRTVLSIASAQGNVEVVRTLLDRGLDENHR 664

Qy   538 DYKGFTPLHPVAGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENI 597
              |:  || |              |  |: |||:      ::| ::::  | :| ::
Db   665 DDAGWTPLHMAAFEGHR------------------LICEALIEQGARTNEIDNDGR--- 702

Qy   598 LGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGY--NAMQ 654
              :                  ::  ::    ::                  ||::||
Db   703 -------IPFILASQEGHYDCVQILLENKSNI 739

Qy   655 VAALFGDKEAVKMLAKSAKPSDLNFK-TSATP 685
              ||| || |   |||     ||
Db   740 VAALEGHRDIVELLFSHG--ADVNCKDADGRPTLYILALENQLTMAEYFLENGANVEASD 797

Qy   686 ----TPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDKNTVLHYAISKGDSFLVQKILAH 741
              :|:|||:||:|||: ||||| ||   : :  | |:    :|::||:|
Db   798 AEGRTALHVSCWQGHMEMVQVLIAYHA-DVN-AADNEKRSALQSAAWQGHVKVVQLLIEH 855

Qy   742 TG-VDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKE 800
              || :||:   ::|:||||||||||:||| | | ||     |: : : :||::||||||
Db   856 GAVVDHTC--NQGATALCIAAQEGHIDVVQVLLEHGADPNHADQGFGRTAMRVAAKNGHSQ 913

Qy   801 KGVLGIVNKLLDRGADINLDG 821
              |  :|:||||:||||||  |
Db   914 ------IIKLLEKYGASSLNG 928
```

```
             GenCore version 5.1.6
    Copyright (c) 1993 - 2005  Compugen Ltd.

OM protein - protein search, using sw model

Run on:    February 26, 2005, 18:45:13 ; Search time 50 Seconds
                                  (without alignments)
                             1599.123 Million cell updates/sec

Title:           US-10-701-038-2
Perfect score:   4310
Sequence:        1 PSGDIQDQSQDQQQBQDQQQ.........DRGADINLDGDRNILFDQCL 831

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters:     283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  PIR_79:*
            1: pir1:*
            2: pir2:*
            3: pir3:*
            4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	391.5	9.1	1856	2	B35049	ankyrin 1, erythro
2	391.5	9.1	1880	2	A35049	ankyrin 1, erythro
3	390.5	9.1	1881	1	SJHUK	ankyrin 1, erythro
4	382.5	8.9	1848	2	S37771	ankyrin, erythrocy
5	381.5	8.9	1862	2	I49502	ankyrin - mouse
6	378.5	8.8	1765	2	T42714	ankyrin 3, splice
7	378.5	8.8	1940	2	T42715	ankyrin 3, splice
8	378.5	8.8	1943	2	T42716	ankyrin 3, splice
9	378.5	8.8	1961	2	T42716	ankyrin 3, splice
10	377.5	8.8	4377	2	A55575	ankyrin 2, long sp
11	368.5	8.5	3924	2	S37431	ankyrin 2, neurona
12	345	8.0	791	2	T42691	hypothetical prote
13	344	8.0	1401	2	S11527	alpha-latrotoxin p
14	342	7.9	934	1	H71274	probable ankyrin -
15	339.5	7.9	1549	2	T13940	ankyrin - fruit fl
16	334.5	7.8	2039	2	T15347	ankyrin-related un
17	321	7.4	1411	2	S30355	alpha-latroinsecto
18	313.5	7.3	1062	2	T14151	Inv protein - mous
19	313.5	7.3	1062	2	T30255	inversin - mouse
20	269.5	6.3	1184	2	T00253	gene Ankhzn protei
21	269	6.2	397	2	T46445	hypothetical prote
22	263	6.1	1188	2	T19552	hypothetical prote
23	255	5.9	426	2	AE2149	hypothetical prote
24	254	5.9	633	2	T27499	hypothetical prote
25	250	5.8	907	2	I50404	p50B/p97 (Lyt-10)
26	249	5.8	1031	2	T43458	hypothetical prote
27	227	5.3	1423	1	I37275	death-associated p
28	225.5	5.2	643	2	D86167	protein F21B7.27 [
29	225	5.2	557	2	T46507	hypothetical prote
30	225	5.2	616	2	T00894	hypothetical prote
31	224.5	5.2	627	2	B86190	hypothetical prote
32	221	5.1	662	2	B84725	ankyrin-like prote
33	221	5.1	1107	2	T21280	hypothetical prote
34	218	5.1	984	2	A41996	NF-kappa-B p50 sub
35	215	5.0	991	2	T25412	hypothetical prote
36	214.5	5.0	842	2	T32258	hypothetical prote
37	214.5	5.0	851	2	T12503	hypothetical prote
38	213	4.9	1058	2	D82654	ankyrin-like prote
39	212.5	4.9	968	2	A37867	transcription fact
40	212	4.9	591	2	T48596	ankyrin-like prote
41	210.5	4.9	606	2	AC2508	hypothetical prote
42	210	4.9	2584	2	T24158	hypothetical prote
43	210	4.9	2606	2	T24157	hypothetical prote
44	210	4.9	2703	1	A24420	notch protein - fr
45	209	4.8	2555	2	A40043	notch protein homo

ALIGNMENTS

```
RESULT 1
B35049
ankyrin 1, erythrocyte splice form 3 - human
N;Alternate names: ankyrin 2.1, erythrocyte; ankyrin-R
N;Contains: ankyrin 2.2, erythrocyte
C;Species: Homo sapiens (man)
C;Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 10-Jul-1998
C;Accession: B35049
R;Lambert, S.; Yu, H.; Prchal, J.T.; Lawler, J.; Ruff, P.; Speicher, D.; Cheung, M.C.; ...
Proc. Natl. Acad. Sci. U.S.A. 87, 1730-1734, 1990
A;Title: cDNA sequence for human erythrocyte ankyrin.
A;Reference number: A35049; MUID:90175370; PMID:1689849
A;Accession: B35049
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1856 <LAM>
C;Genetics:
A;Gene: GDB:ANK1; ANK
A;Cross-references: GDB:118737; OMIM:182900
A;Map position: 8p11.2-8p11.2
C;Superfamily: ankyrin; ankyrin repeat homology
C;Keywords: alternative splicing
F;2-1856/Product: ankyrin 1, erythrocyte form 3 #status predicted <MAT>
F;2-1513,1676-1856/Product: ankyrin 2.2, erythrocyte #status predicted <MA2>
F;44-76/Domain: ankyrin repeat homology <AN01>
F;77-109/Domain: ankyrin repeat homology <AN02>
F;110-142/Domain: ankyrin repeat homology <AN03>
F;143-171/Domain: ankyrin repeat homology <AN04>
F;172-204/Domain: ankyrin repeat homology <AN05>
F;205-237/Domain: ankyrin repeat homology <AN06>
F;238-270/Domain: ankyrin repeat homology <AN07>
F;271-303/Domain: ankyrin repeat homology <AN08>
F;304-336/Domain: ankyrin repeat homology <AN09>
F;337-369/Domain: ankyrin repeat homology <AN10>
F;370-402/Domain: ankyrin repeat homology <AN11>
F;403-435/Domain: ankyrin repeat homology <AN12>
F;436-468/Domain: ankyrin repeat homology <AN13>
F;469-501/Domain: ankyrin repeat homology <AN14>
F;502-534/Domain: ankyrin repeat homology <AN15>
F;535-567/Domain: ankyrin repeat homology <AN16>
F;568-600/Domain: ankyrin repeat homology <AN17>
F;601-633/Domain: ankyrin repeat homology <AN18>
F;634-666/Domain: ankyrin repeat homology <AN19>
F;667-699/Domain: ankyrin repeat homology <AN20>
F;700-732/Domain: ankyrin repeat homology <AN21>
F;733-765/Domain: ankyrin repeat homology <AN22>
F;766-798/Domain: ankyrin repeat homology <AN23>

Query Match             9.1%;  Score 391.5;  DB 2;  Length 1856;
Best Local Similarity   26.0%;  Pred. No. 1.2e-14;
Matches  174;  Conservative  100;  Mismatches  246;  Indels  149;  Gaps  31;
```

```
F;2-1513,1676-1880/Product: ankyrin 2.2, erythrocyte #status predicted <MA2>
F;44-76/Domain: ankyrin repeat homology <AN01>
F;77-109/Domain: ankyrin repeat homology <AN02>
F;110-142/Domain: ankyrin repeat homology <AN03>
F;143-171/Domain: ankyrin repeat homology <AN04>
F;172-204/Domain: ankyrin repeat homology <AN05>
F;205-237/Domain: ankyrin repeat homology <AN06>
F;238-270/Domain: ankyrin repeat homology <AN07>
F;271-303/Domain: ankyrin repeat homology <AN08>
F;304-336/Domain: ankyrin repeat homology <AN09>
F;337-369/Domain: ankyrin repeat homology <AN10>
F;370-402/Domain: ankyrin repeat homology <AN11>
F;403-435/Domain: ankyrin repeat homology <AN12>
F;436-468/Domain: ankyrin repeat homology <AN13>
F;469-501/Domain: ankyrin repeat homology <AN14>
F;502-534/Domain: ankyrin repeat homology <AN15>
F;535-567/Domain: ankyrin repeat homology <AN16>
F;568-600/Domain: ankyrin repeat homology <AN17>
F;601-633/Domain: ankyrin repeat homology <AN18>
F;634-666/Domain: ankyrin repeat homology <AN19>
F;667-699/Domain: ankyrin repeat homology <AN20>
F;700-732/Domain: ankyrin repeat homology <AN21>
F;733-765/Domain: ankyrin repeat homology <AN22>
F;766-798/Domain: ankyrin repeat homology <AN23>
```

```
Query Match             9.1%;   Score 391.5;  DB 2;  Length 1880;
Best Local Similarity   26.0%;  Pred. No. 1.2e-14;
Matches  174;  Conservative 100;  Mismatches 246;  Indels 149;  Gaps 31;

Qy    194 RIVASKGDLSEDQVEEILDIIFMNESEIAEGISNPLHADVDNNPVK-------GA---- 241
          |:: :|     : ::::                            |:
Db    256 RLLLDRGAQIETKTKDEL-----------------TPLHCAARNGHVRISEILLDHGAPIQA 300

Qy    242 --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT 298
          :|: :|:| : : :                           ||||:
Db    301 KTKNGLSPIHMAAQGDHLD--CVRLLLQYDAEIDDI------TLD--HLTPLHVAAHCG 349

Qy    299 GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK 357
          |||  :::|                    |||||:
Db    350 HHRVAKVLLDKGAKPNSRALN-GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL 405

Qy    358 NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYF---GNT-IAHYAVKNK 407
          ||| :   ::|:||                         :|:  :| :|:|
Db    406 TPLHVASFMGHLPIVKNLLQRGASPNVSNVK------VETPLHMAARAGHTEVAKYLLQNK 460

Qy    408 NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP 467
                     :|:                           |||||:
Db    461 ------------AKVNAKADDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP 506

Qy    468 LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK 525
          |||| |  | ::|| :|| :|   |:: | ||||:::||:||: |::|
Db    507 LHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA 565

Qy    526 GEDTVAKSLLMEDYKGFTPLHFVAGGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE 585
          | :         ||||||:
Db    566 GKN-----------------GLTPLHVAVHHNNLDIVKL------------------- 587

Qy    586 LVNLGDFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP 645
          |:        | |:      |:                        |:|::||:
Db    588 LLPRG-------GSP------HSPAWN-GYTPLHIAAKQNQVEVARSLLQY-GGSANAE 631

Qy    646 SERGYNAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG 705
          |::|  | | :|: :    ||::|            ||||:::|: ||:|:|
Db    632 SVQGVTPLHLAAQEGHAEMVALLL---SKQANGNLGNKSGLTPLHVAQEGHVPVADVLI- 688

Qy    706 QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG 763
          :||    |:    :  |||| | |  : |::|:||::: :||||
Db    689 KHGVMVDATTRMG---YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQ 744

Qy    764 GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL---LDRGADINLD 820
          | :: |: |||:|| ||    :| | |::|     ::||: :|:  :   ||:|| |:
Db    745 GHTDIVTLLLKNGASPNEVSSDGTTPLAIA-----KRLGYISVTDVLKVVTDETSFVLVS 799

Qy    821 GDHNILFDQ 829
          :|:     :
Db    800 DKHRMSFPE 808
```

```
RESULT 2
A35049
ankyrin 1, erythrocyte splice form 2 - human
N;Alternate names: ankyrin 2.1, erythrocyte; ankyrin-R
N;Contains: ankyrin 2.2, erythrocyte
C;Species: Homo sapiens (man)
C;Date: 27-Jul-1990 #sequence_revision 01-Oct-1992 #text_change 09-Jul-2004
C;Accession: A35049
R;Lambert, S.; Yu, H.; Prchal, J.T.; Lawler, J.; Ruff, P.; Speicher, D.; Cheung, M.C.;
Proc. Natl. Acad. Sci. U.S.A. 87, 1730-1734, 1990
A;Title: cDNA sequence for human erythrocyte ankyrin.
A;Reference number: A35049; MUID:90175370; PMID:1689849
A;Accession: A35049
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1880 <LAM>
A;Cross-references: UNIPROT:P16157; GB:M28880
C;Genetics:
A;Gene: GDB:ANK1; ANK
A;Map position: 8p11.2-8p11.2
A;Cross-references: GDB:118737; OMIM:182900
C;Superfamily: ankyrin; ankyrin repeat homology
C;Keywords: alternative splicing; cytoskeleton
F;2-1880/Product: ankyrin 1, erythrocyte form 2 #status predicted <MAT>
```

us-10-701-038-2.rpr

```
Qy    821  GDHNILFDQ 829
                |.|:|
Db    800  DKHRMSFPE 808


RESULT 3
SJHUK
ankyrin 1, erythrocyte splice form 1 - human
N;Alternate names: ankyrin 2.1, erythrocyte; ankyrin-R
N;Contains: ankyrin 2.2
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: S08275; A33219; PC2220; A35443
R;Lux, S.E.; John, K.M.; Bennett, V.
Nature 344, 36-42, 1990
A;Title: Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure w
A;Reference number: S08275; MUID:90158830; PMID:2137557
A;Accession: S08275
A;Molecule type: mRNA
A;Residues: 1-1881 <LU1>
A;Cross-references: UNIPROT:P16157; EMBL:X16609; NID:g28701; PIDN:CAA34610.1; PID:g28702
A;Accession: A33219
A;Molecule type: protein
A;Residues: 2-7,'X',9-17,'X',19-20,'T',22-30,733-749,'A',751-753;828-833,'X',835-855,'X'
X',1367;1383-1427;1601-1630;1686-1698,'D',1700;1763-1772 <LUX>
A;Note: 845-Arg and 1392-Thr were also found
R;Hermann, J.; Barel, H.; Frade, R.
Biochem. Biophys. Res. Commun. 204, 453-460, 1994
A;Title: Human erythrocyte ankyrin, a cytoskeleton component, generates the p57 membrane
A;Reference number: PC2220; MUID:95071348; PMID:7526850
A;Accession: PC2220
A;Molecule type: protein
A;Residues: 910-929 <HER>
R;Davis, L.H.; Bennett, V.
J. Biol. Chem. 265, 10589-10596, 1990
A;Title: Mapping the binding sites of human erythrocyte ankyrin for the anion exchanger
A;Reference number: A35443; MUID:90285190; PMID:2141335
A;Accession: A35443
A;Molecule type: protein
A;Residues: 'X',5,'X',7-12;403-417,'X',419-422,'H',424,'LQ',797-800,'L',802-814;862-863,
C;Genetics:
A;Gene: GDB:ANK1; ANK
A;Cross-references: GDB:118737; OMIM:182900
A;Map position: 8p11.2-8p11.2
C;Superfamily: ankyrin; ankyrin repeat homology
C;Keywords: alternative splicing; phosphoprotein
F;2-1881/Product: ankyrin 1, erythrocyte form 1 #status predicted <MAT1>
F;2-1512,1675-1881/Product: ankyrin 2.2, erythrocyte #status predicted <MAT2>
F;2-827/Domain: 89K #status predicted <DOM1>
F;2-827/Region: anion exchange protein binding <AN01>
F;44-76/Domain: ankyrin repeat homology <AN01>
F;77-109/Domain: ankyrin repeat homology <AN02>
F;110-142/Domain: ankyrin repeat homology <AN03>
F;143-171/Domain: ankyrin repeat homology <AN04>
F;172-204/Domain: ankyrin repeat homology <AN05>
F;205-237/Domain: ankyrin repeat homology <AN06>
F;238-270/Domain: ankyrin repeat homology <AN07>
F;271-303/Domain: ankyrin repeat homology <AN08>
F;304-336/Domain: ankyrin repeat homology <AN09>
F;337-369/Domain: ankyrin repeat homology <AN10>
F;370-402/Domain: ankyrin repeat homology <AN11>
F;403-435/Domain: ankyrin repeat homology <AN12>
F;436-468/Domain: ankyrin repeat homology <AN13>
F;469-501/Domain: ankyrin repeat homology <AN14>
F;502-534/Domain: ankyrin repeat homology <AN15>
F;535-567/Domain: ankyrin repeat homology <AN16>
F;568-600/Domain: ankyrin repeat homology <AN17>
F;601-633/Domain: ankyrin repeat homology <AN18>
F;634-666/Domain: ankyrin repeat homology <AN19>
F;667-699/Domain: ankyrin repeat homology <AN20>
F;700-732/Domain: ankyrin repeat homology <AN21>
F;733-765/Domain: ankyrin repeat homology <AN22>
F;766-798/Domain: ankyrin repeat homology <AN23>
F;828-1382/Domain: 62K #status predicted <DOM2>
F;828-1382/Region: spectrin binding
F;1383-1881/Domain: 55K #status predicted <DOM3>


Query Match            9.1%;  Score 390.5; DB 1;  Length 1881;
Best Local Similarity  26.0%; Pred. No. 1.4e-14;
Matches 174;  Conservative 100;  Mismatches 246;  Indels 149;  Gaps 31;


Qy    194  RIVASKGDLSEDQVEBILDIIFMNESEIAEGISNPLHADVDNNPVK---------GA--- 241
                |::| .:| .: .:.|.:                           |||          |
Db    256  RLLLDRGAQIETKTKDEL---------------TPLHCAARNGHVRISEILLDHGAPIQA 300

Qy    242  --KNVMTLMHLVYACD-VDPRIVKALGEVENDEGDLGANAYNVLDSEGNLPLHHAAKNCT 298
             ||  :|.:||.:   |||.   .: |             ||| :|| :|.|.|
Db    301  KTKNGLSPIHMAAQGDHLD--CVRLLQYDAEIDDI------TLD--HLTPLHVAAHCG 349

Qy    299  GDKL-KLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEEFTSRNLDFGLVDGDGK 357
             |||  .                       :|||||||  .|:|||||| .||||:||||
Db    350  HHRVAKVLLDKGAKPNSRALN--GFTPLHIACKKNHVRVMELLLKTGASID--AVTESGL 405

Qy    358  NPLHHA--VEHLPPV---ILKGVMDHVKNSSEFQDLVNDPDYP----GNT-IAHYAVKNK 407
             |||||.                .:|  : .: .| :                ::||
Db    406  TPLHVASFMGHLPIVKNLLQRGASPNVSNVK-----VETPLHMAARAGHTEVAKYLLQNK 460

Qy    408  NADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTP 467
                       : | :|.|||  .| : .|.||. .:|       :|.   |  |.
Db    461  ------------AKVNAKADDQTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTP 506

Qy    468  LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNIK 525
             |||||  :||  ||. |||||.|..:.:|||.||. ||||||||:.||      ||||:|
Db    507  LHIAAREGHVETVLALL-EKEASQACMTKKGFTPLHVAAKYGKVRVAELLLERDAHPNAA 565

Qy    526  GEDTVAKSLLMEDYKGFTPLHFVAGGSGRDTFRVVRKNYEKCHDLATIRAALMQDRSGGE 585
             :                   |||:                  ||. :.
Db    566  GKN------------GLTPLHVAVHHNNLDIVKL-------------------------- 587

Qy    586  LVNLGDPESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLP 645
             | | ||.||..            .||         .| :|:|  :|
Db    588  LLPRG------GSP------HSPAWN-GYTPLHIAAKQNQVEVARSLLQY-GGSANAE 631

Qy    646  SERGYNAMQVAALFGDKEAVRKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVG 705
             | |:  .|.  |  |:|   :|  .||.        .|  |  | |
Db    632  SVQGVTPLHLRAQEGHAEMVALLL--SKQANGNLGNKSGLTPLHLVAQEGHVPVADVLI- 688

Qy    706  QHG--IDINQRMGSDKNTVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEG 763
                                .|| :|||||||.||.|||.||.|||. |  || ||||: |
Db    689  KHGVMVDATTRMG---YTPLHVASHYGNIKLVKFLLQHQ-ADVNAKTLGYSPLHQAQQ 744

Qy    764  GDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKL--LDRGADINLD 820
             ||||::| ||| ||.||.||:||.|..||.|   |.|      :::.    ||||.|:|
Db    745  GHTDIVTLLLKNGASPNEVSSDGTTPLAIA----KRLGYISVTDVLKVVTDETSFVLVS 799

Qy    821  GDHNILFDQ 829
                |.|:|
Db    800  DKHRMSFPE 808
```

RESULT 4
S37771
ankyrin, erythrocyte - mouse
C;Species: Mus musculus (house mouse)
C;Date: 27-May-1994 #sequence_revision 03-Aug-1995 #text_change 09-Jul-2004
C;Accession: S37771
R;Birkenmeier, C.S.; White, R.A.; Peters, L.L.; Hall, E.J.; Lux, S.E.; Barker, J.E.
J. Biol. Chem. 268, 9533-9540, 1993
A;Title: Complex patterns of sequence variation and multiple 5' and 3' ends are found a
A;Reference number: S37771; MUID:93252825; PMID:8486643
A;Accession: S37771
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1848 <BIR>
A;Cross-references: UNIPROT:Q61302; EMBL:X69063; NID:g311816; PIDN:CAA48801.1; PID:g311
C;Superfamily: ankyrin; ankyrin repeat homology

```
Qy   783  DDNGRSVLSSAIVPGRKEKGVLGIVNKL 810
          ||| |::: :: ||         :::
Db   768  SSNGTTPLAIA------KRLGYISVTDVL 790
```

```
RESULT 5
I49502
ankyrin - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C;Accession: I49502
R;White, R.A.; Birkenmeier, C.S.; Peters, L.L.; Barker, J.E.; Lux, S.E.
Mamm. Genome 3, 281-285, 1992
A;Title: Murine erythrocyte ankyrin cDNA: Highly conserved regions of the regulatory do
A;Reference number: I49502; MUID:92345717; PMID:1386265
A;Accession: I49502
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1862 <RES>
A;Cross-references: UNIPROT:Q02357; GB:M84756; NID:g191939; PID:g191940
C;Genetics:
A;Gene: Ank-1
C;Superfamily: ankyrin; ankyrin repeat homology <AN01>
C;Keywords: alternative splicing
F;40-72/Domain: ankyrin repeat homology <AN01>
F;73-105/Domain: ankyrin repeat homology <AN02>
F;106-138/Domain: ankyrin repeat homology <AN03>
F;139-167/Domain: ankyrin repeat homology <AN04>
F;168-200/Domain: ankyrin repeat homology <AN05>
F;201-233/Domain: ankyrin repeat homology <AN06>
F;234-266/Domain: ankyrin repeat homology <AN07>
F;267-299/Domain: ankyrin repeat homology <AN08>
F;300-332/Domain: ankyrin repeat homology <AN09>
F;333-365/Domain: ankyrin repeat homology <AN10>
F;366-398/Domain: ankyrin repeat homology <AN11>
F;399-431/Domain: ankyrin repeat homology <AN12>
F;432-464/Domain: ankyrin repeat homology <AN13>
F;465-497/Domain: ankyrin repeat homology <AN14>
F;498-530/Domain: ankyrin repeat homology <AN15>
F;531-563/Domain: ankyrin repeat homology <AN16>
F;564-596/Domain: ankyrin repeat homology <AN17>
F;597-629/Domain: ankyrin repeat homology <AN18>
F;630-662/Domain: ankyrin repeat homology <AN19>
F;663-695/Domain: ankyrin repeat homology <AN20>
F;696-728/Domain: ankyrin repeat homology <AN21>
F;729-761/Domain: ankyrin repeat homology <AN22>
F;762-794/Domain: ankyrin repeat homology <AN23>

Query Match             8.9%;   Score 381.5;  DB 2;   Length 1862;
Best Local Similarity   26.0%;  Pred. No. 4.6e-14;
Matches 179;  Conservative 101;  Mismatches 257;  Indels 151;  Gaps 33;

Qy  221  IAEGISNPLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYN 280
Db  148  LQQGHENVV-AHLINYGTKG-KVRLPALHIAARND-DTRTAAVL--LQNDPNP------D 196
Qy  281  VLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQ---SPLHIITQKPDCSVL 337
Db  197  VLSKTGFTPLHIAAHYENLNVAQLLNRGA------SVNFTPQNGITPLHIASRRGNVIMV 251
Qy  338  DI------------SRN----------LDFG----LVDGDGKNPLHHA--- 363
Db  252  RLLLDRGAQIETRTKDELTPLHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHMAAQG 311
Qy  364  ------------------------------VEHLPPVILKGVMDHVKNSSEFQDLVNDPD---YFGNTIA 400
Db  312  DHLDCVRLLLQYNABIDDITLDHLTPLHVAAHCGHHRVAKVLLDKGAKPNSRALNGFTPL 371
Qy  401  HYAVNKNADLTLFMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQ 460
Db  372  HIACKKNH--IRVMELLLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVS 429
Qy  461  DVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--G 518
```

```
Query Match             8.9%;   Score 382.5;  DB 2;   Length 1848;
Best Local Similarity   26.2%;  Pred. No. 4e-14;
Matches 180;  Conservative 98;  Mismatches 259;  Indels 151;  Gaps 33;

Qy  221  IAEGISNPLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYN 280
Db  156  LQQGHENVV-AHLINYGTKG-KVRLPALHIAARND-DTRTAAVL--LQNDPNP------D 204
Qy  281  VLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQ---SPLHIITQKPDCSVL 337
Db  205  VLSKTGFTPLHIAAHYENLNVAQLLNRGA------SVNFTPQNGITPLHIASRRGNVIMV 259
Qy  338  DI------------SRN----------LDFG----LVDGDGKNPLHHA--- 363
Db  260  RLLLDRGAQIETRTKDELTPLHCAARNGHVRISEILLDHGAPIQAKTKNGLSPIHMAAQG 319
Qy  364  ------------------------------VEHLPPVILKGVMDHVKNSSEFQDLVNDPD---YFGNTIA 400
Db  320  DHLDCVRLLLQYNABIDDITLDHLTPLHVAAHCGHHRVAKVLLDKGAKPNSRALNGFTPL 379
Qy  401  HYAVNKNADLTLFMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQ 460
Db  380  HIACKKNH--IRVMELLLKTGASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVS 437
Qy  461  DVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--G 518
Db  438  NVKVETPLHMAAARGHTE-VAKYLLQNKAKANAKADDQTPLHCAARIGHTGMVKLLLEN  496
Qy  519  SAKPNI--------------------KGEDTVAKSLLMED----------YKGFTPLHPVAGGGSRDTF 557
Db  497  GASPNLATTAGHTPLHTAAREGHVDTALALLEKEASQACMTKKGFTPLHVAAKYGK---V 553
Qy  558  RVVRKNYEKCHD---------LATIRAALMQDRSGGELVNLGDFESENIL---GSPNAK 604
Db  554  RLABLLLE--HDAHPNAAGKNGLTPLHVAVHHNNL--DIVKL------LLPRCGSP--- 599
Qy  605  FLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEA 664
Db  600  ---HSPAWN-GYTPLHAAKQNQIEVARSLLQY-GGSANAESVQGVTPLHLAAQEGHTEM 654
Qy  665  VKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHG---IDINQRMGSDKNTV 722
Db  655  VALLL--SKQANGNLGNKSGLTPLHLVSQEGHVPVADVLI-KHGVTVDATTRMG---YTP 708
Qy  723  LHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRL 782
Db  709  LHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQGHTDIVTLLLKNGASPNEV 767
```

Db   430 NVKVETPLHMAARAGHTE-VAXYLLQNKAKANAKADDQTPLHLHCAARIGHTGMVKLLLEN 488
         : | |||||||| | || |||| |
Qy   389 IACK-KNR-IRVMELLLKHGASIQAVTESGLTIPHVAAFMGHVNIVSQLMHHGASPNTTN 446

Qy   519 SAKPNI-----------KGEDTVAKSLLMED-------YKGFTPLHFVAGGGSRDTF 557
         : |||| ||||||| |||| ||||||||||||||| |
Db   462 VNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GS 519

Db   489 GASPNLATTAGHTPLHTAAREGHDTALALLEKEASQACMTKKGFTPLHVAAKYGK---V 545
           ||||| | |||||| || || || |||| |||| |||||||||||| :
Qy   447 VRGETALHMAARSGQAE-VVRYLVQDGAQVEAKAKDQTPLHISARLGKADIVQQLLQQG 505

Qy   558 RVVRKNYEKCHD------------LATIRAALMQDRSGGELVNLGDFESENIL--GSPNAK 604
         : |||| ||| ||| ||||| ||| | || : |||||
Db   520 AKPNI------------KGEDTVA-------KSLLMEDYKGFTPLHFVAGGGSRDTFR 558

Db   546 RLAELLE--HDAHPNAAGKNGLTPLHVAVHHNNL--DIVKL------LLPRGGSP--- 591
          | | | | ||| ||||||| || | | || || |
Qy   506 ASPNAATTSGYTPLHLAAREGHEDVAAPLLDHGASLSITTKKGFTPLHVAAKYGKLEV-- 563

Qy   605 FLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEA 664
         ||||| ||||| | || || | ||| | ||||||||| || | ||| |
Db   559 VVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGSNAKFLQHIQSANFGFSP 618

Db   592 ---HSPAWN-GYTPLHIAAKQNQIEVARSLLQY-GGSANAESVQGVTPLHLHAAQEGHTEM 646
            : | | ||| ||| || ||| | |||
Qy   564 ASLLQKS----------------------ASPDA---------AGKSGLTP 584

Qy   665 VKMLAKSAKPSDLNFKTSATPPLNLACLRGDNEVVRGLVGQHG--IDINQRMGSDKNTV 722
         | ||| |||| | |||| |||||||||||| ||| || ||||| |||| |
Db   619 ARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEAVVMLAKSAKPSDLN 678

Db   647 VALLL--SKQANGLGNKSGLTPLHLVSQEG-HVLVADVLIKHGVTVDATTRMG---YTP 700
          :|| | ||| ||||||| | | | | | | ||
Qy   679 LHVAAHYDNQKVALLLD-QGASPHAAAKNGYTPLHIAAKKNQMDIATSLLEVG--ADAN 737

Qy   723 LHYAISKGDSFLVQKIIAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRL 782
         ||| ||||||| ||| |||||||||||||||||||||| ||||||||||||||||||
Db   701 LHVASHYGNIKLVKFLLQHQ-ADVNAKTKLGYSPLHQAAQQGHTDIVTLLLKNGASPNEV 759

Qy   783 DDNGRSVLSSAIVPGRKEKGVLGIVNKL 810
         : ||||||||||| |||||||||
Db   760 SSNGTTPLAIA-----KRLGYISVTDVL 782

Qy   798 RKEKGVLGIVNKLLDRGADIN 818
                  ||| |||| | |
Db   754 -AQQGHTHIINVLLQNNASPN 773

RESULT 6
T42714
ankyrin 3, splice form 2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C;Accession: T42714
R;Peters, L.L.; John, K.M.; Lu, F.M.; Eicher, E.M.; Higgins, A.; Yialamas, M.; Turtzo, L
J. Cell Biol. 130, 313-330, 1995
A;Title: Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene
the repeat domain.
A;Reference number: Z22237; MUID:95340633; PMID:7615634
A;Accession: T42714
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1765 <PET>
A;Cross-referendes: UNIPROT:Q61307; EMBL:L40632; NID:g710548; PID:g710551; PIDN:AAB01605
A;Experimental source: strain C57BL/6J; kidney
C;Genetics:
A;Gene: Ank3
A;Map position: 10
A;Introns: 1587/1   Indels   133;   Gaps   27;
C;Superfamily: ankyrin; ankyrin repeat homology
C;Keywords: alternative splicing

Query Match   8.8%;   Score 378.5;   DB 2;   Length 1765;
Best Local Similarity 25.7%;   Pred. No. 6.4e-14;
Matches 175; ; Conservative   98;   Mismatches 275;   Indels 133;   Gaps 27;

Qy   184 TSGPVRDIADRIVASKGDLSEDQVEBILDIIFMNES---BIAEGISNPLH----------- 230
         |||||||| ||||||||| | ||||| ||||||| |: |:||||||
Db   180 TKGKVRLPALHIAARKDD------TKAAALLQNDTNADVESKSGFT-PLHIAAHYGNIN 232

Qy   231 ------------ADVDNNPVKGAKNVMTLMHLVTACDVDPRIVASKGNAMVKL-------- 281
                     | ||| || ||||| |||| |||| ||| |
Db   233 VATLLLNRAAAVDFT---ARNDITPLH-VASKRGNAMVKLL------LDRGAK-IDA 279

Qy   282 LDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEE 341
         |||||||||||| || | |||| ||||||| ||||||||||| | |||
Db   280 KTRDGLTPLHCGARSGHEQVVEMLLDRSAPILSKTKN--GLSPLHMATQGDHLNCVQL-- 335

Qy   342 FTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAH 401
         |||||||||| ||||||||| |||||| ||||||||||| | |||||||| |||| | |
Db   336 LLQHNVPVDDVTNDYLTALHVAAHCGHYKAVAKVLLD--KKASPNAKALN-----GFTPLH 388

Qy   402 YAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQD 461

RESULT 7
T42715
ankyrin 3, splice form 3 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C;Accession: T42715
R;Peters, L.L.; John, K.M.; Lu, F.M.; Eicher, E.M.; Higgins, A.; Yialamas, M.; Turtzo, L
J. Cell Biol. 130, 313-330, 1995
A;Title: Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene
the repeat domain.
A;Reference number: Z22237; MUID:95340633; PMID:7615634
A;Accession: T42715
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1940 <PET>
A;Cross-references: UNIPROT:Q61307; EMBL:L40632; NID:g710548; PID:g710549; PIDN:AAB0160
A;Experimental source: strain C57BL/6J; kidney
C;Genetics:
C;Gene: Ank3
A;Map position: 10
A;Introns: 834/1
C;Superfamily: ankyrin; ankyrin repeat homology
C;Keywords: alternative splicing

Query Match   8.8%;   Score 378.5;   DB 2;   Length 1940;
Best Local Similarity 25.7%;   Pred. No. 7.5e-14;
Matches 175;   Conservative   98;   Mismatches 275;   Indels 133;   Gaps 27;

Qy   184 TSGPVRDIADRIVASKGDLSEDQVEBILDIIFMNES---BIAEGISNPLH----------- 230
         |||||||| ||||||||| | ||||| ||||||| |: |:||||||
Db   180 TKGKVRLPALHIAARKDD------TKAAALLQNDTNADVESKSGFT-PLHIAAHYGNIN 232

Qy   231 ------------ADVDNNPVKGAKNVMTLMHLVTACDVDPRIVALGEVENDEGDLGANAYNV 281
                     | ||| || ||||| |||| |||| || | | |
Db   233 VATLLLNRAAAVDFT---ARNDITPLH-VASKRGNAMVKLL------LDRGAK-IDA 279

Qy   282 LDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDCSVLDIEE 341
         |||||||||||| || | |||| ||||||| ||||||||||| | |||
Db   280 KTRDGLTPLHCGARSGHEQVVEMLLDRSAPILSKTKN--GLSPLHMATQGDHLNCVQL-- 335

Qy   342 FTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPYFGNTIAH 401

```
Db  233 VATLLLNRAAAVDFT-----ARNDITPLH-VASKRGNANMVKLL------LDRGAK-IDA 279
                  :|    |||                              |:
Qy  282 LDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFQSPLHIITQKPDCSVLDIEE 341
        |     ||:  :    |||                    |::
Db  280 KTRDGLTPLHCGARSGHEQVVEMLLDRSAPILSKTKN--GLSPLHMATQGDHLNCVQL-- 335
          | |  ||                              ||
Qy  342 FTSRNLDFGLVDGGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPYFGNTIAH 401
             :               |                                :
Db  336 LLQHNVPVDDVTNDYLTALHVAAHCGHYKVAKVLLD--KKASPNAKALN------GFTPLH 388
         |:                |||                              |||
Qy  402 YAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQD 461
         :|    |:    |||                         ::|
Db  389 IACK-KNR-IRVMELLLKHGASIQAVTESGLTPIHVAAFMGHVNIVSQLMHHGASPNTTN 446
            |    ||                  ||||:   :
Qy  462 VNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKAL--GS 519
         |||||                          :|||::   :
Db  447 VRGETALHMAARSGQAE-VVRYLVQDGAQVEAKAKDDQTPLHISARLGKADIVQQLLQQG 505
          | |||                          |||:   :
Qy  520 AKPNI--------------KGEDTVA----------KSLLMEDYKGFTPLHFVAGGSRDTFR 558
         |  |              |||:|:              :::::|||  ||||
Db  506 ASPNAATTSGYTPLHLAAREGHEDVAAFLLDHGASLSITTKKGFTPLHVAAKYGKLEV-- 563
                  ||||:      |:  |                  :||||:
Qy  559 VVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSP 618
                                  |:| :
Db  564 -----------------ASLLLQKS----------ASPDA------AGKSGLTP 584
Qy  619 ARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAKSAKPSDLN 678
             ||          |:|:   :|:|:    :
Db  585 LHVAAHYDNQKVALLLLD-QGASPHAAAKNGYTPLHIAKKNQMDIATSLLEYG--ADAN 641
            ::|   |:|                   ||||:
Qy  679 FKTSATPTPLNLACLRGDNEVVRGLVGQH-GIDINQRMGSDKNTVLHYAISKGDSFLVQK 737
          ||||||||                     |||:         :||||
Db  642 AVTRQGIASVHLAAQEGHVDMVSLLLSRNANVNLSNKSGL---TPLHLA-AQEDRVNVAE 697
                  ||:||              :  |:| ||    |||| :
Qy  738 ILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPG 797
          |:| |||                 |:||||            |:|:
Db  698 VLVNQGAHVDAQTKMGYTPLHVGCHYGNIKIVNFLLQHSAKVNAKTKNGYTALHQA----- 753
          |:|||| ||             ||| |:       |:||
Qy  798 RKEKGVLGIVNKLLDRGADIN 818
         :||  |||  |||
Db  754 -AQQGHTHIINVLLQNNASPN 773
```

RESULT 9
T42716
ankyrin 3, splice form 4 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C;Accession: T42716
R;Peters, L.L.; John, K.M.; Lu, F.M.; Eicher, E.M.; Higgins, A.; Yialamas, M.; Turtzo, L
J. Cell Biol. 130, 313-330, 1995
A;Title: Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene
the repeat domain.
A;Reference number: Z22237; MUID:95340633; PMID:7615634
A;Accession: T42716
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1961 <PET>
A;Cross-references: UNIPROT:Q61307; EMBL:L40632; NID:g710548; PID:g710552; PIDN:AAB01607
A;Experimental source: strain C57BL/6J; kidney
C;Genetics:
A;Gene: Ank3
A;Map position: 10
C;Superfamily: ankyrin; ankyrin repeat homology
C;Keywords: alternative splicing

```
Query Match       8.8%;   Score 378.5;   DB 2;   Length 1961;
Best Local Similarity   25.7%;   Pred. No. 7.6e-14;
Matches 175;   Conservative  98;   Mismatches 275;   Indels 133;   Gaps 27;

Qy  184 TSGPVRDIADRIVASKGDLSEDQVEBILDIIFMNES----EIAEGISNPLH--------- 230
              :| |             ::|:                  :|||
Db  180 TKGKVRLPALHIAARKDD------TKAAALLQNDTNADVESKSGFT-PLHIAAHYGNIN 232
```

RESULT 8
T42713
ankyrin 3, splice form 1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C;Accession: T42713
R;Peters, L.L.; John, K.M.; Lu, F.M.; Eicher, E.M.; Higgins, A.; Yialamas, M.; Turtzo, L
J. Cell Biol. 130, 313-330, 1995
A;Title: Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene
the repeat domain.
A;Reference number: Z22237; MUID:95340633; PMID:7615634
A;Accession: T42713
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1943 <PET>
A;Cross-references: UNIPROT:Q61307; EMBL:L40632; NID:g710548; PID:g710550; PIDN:AAB01606
A;Experimental source: strain C57BL/6J; kidney
C;Genetics:
A;Gene: Ank3
A;Map position: 10
A;Introns: 855/1
C;Function:
A;Description: supposed to play an important role in the polarized distribution of many
A;Note: major kidney ankyrin
C;Superfamily: ankyrin; ankyrin repeat homology
C;Keywords: alternative splicing

```
Query Match       8.8%;   Score 378.5;   DB 2;   Length 1943;
Best Local Similarity   25.7%;   Pred. No. 7.5e-14;
Matches 175;   Conservative  98;   Mismatches 275;   Indels 133;   Gaps 27;

Qy  184 TSGPVRDIADRIVASKGDLSEDQVEBILDIIFMNES----EIAEGISNPLH--------- 230
              :| |             ::|:                  :|||
Db  180 TKGKVRLPALHIAARKDD------TKAAALLQNDTNADVESKSGFT-PLHIAAHYGNIN 232
Qy  231 ---------ADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNV 281
```

```
F;234-266/Domain: ankyrin repeat homology <AN06>
F;267-299/Domain: ankyrin repeat homology <AN07>
F;300-332/Domain: ankyrin repeat homology <AN08>
F;333-365/Domain: ankyrin repeat homology <AN09>
F;366-398/Domain: ankyrin repeat homology <AN10>
F;399-431/Domain: ankyrin repeat homology <AN11>
F;432-464/Domain: ankyrin repeat homology <AN12>
F;465-497/Domain: ankyrin repeat homology <AN13>
F;498-530/Domain: ankyrin repeat homology <AN14>
F;531-563/Domain: ankyrin repeat homology <AN15>
F;564-596/Domain: ankyrin repeat homology <AN16>
F;597-629/Domain: ankyrin repeat homology <AN17>
F;630-662/Domain: ankyrin repeat homology <AN18>
F;663-695/Domain: ankyrin repeat homology <AN19>
F;696-728/Domain: ankyrin repeat homology <AN20>
F;729-761/Domain: ankyrin repeat homology <AN21>
F;762-794/Domain: ankyrin repeat homology <AN22>
F;795-827/Domain: ankyrin repeat homology <AN23>

Query Match          8.8%;   Score 377.5;  DB 2;  Length 4377;
Best Local Similarity   25.5%;   Pred. No. 3.2e-13;
Matches  175;  Conservative  92;  Mismatches 277;  Indels 141;  Gaps 26;

Qy   184 TSGPVRDIADRIVASKGDLSEDQVEEILDIIFMNES------EIAEGISNPLH----- 230
Db   197 TKGKVRLPALHIAARKDD--------TKAAALLQNDNADVESKSGFT-PLHIAAHYGNIN 249

Qy   231 -------ADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNV 281
Db   250 VATLLLNRAAAVDFT---ARNDITPLH-VASKRGNANMVKLL------LDRGAK-IDA 296

Qy   282 LDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFQSPLHIITQKPDCSVLDIEE 341
Db   297 KTRDGLTPLHCGARSGHEQVVEMLLDRAAPILSKTKN--GLSPLHMATQGDHLNCVQL-- 352

Qy   342 FTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAH 401
Db   353 LLQHNVPVDDVTNDYLTALHVAAHCGHYKVAKVLLDKKANP------NAKALNGFTPLH 405

Qy   402 YAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQD 461
Db   406 IACK-KNR-IKVMELLLKHGASIQAVTESGLTPIHVAAFMGHVNIVSQLMHHGASPNTTN 463

Qy   462 VNGDTPLHIAVEBGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIVGIKAL--GS 519
Db   464 VRGETALHMAARSGQAE-VVRYLVQDGAQVEAKAKDDQTPLHISARLGKADIVQQLLQQG 522

Qy   520 AKPNI-----------------KGEDTVA------KSLLMEDYKGFTPLHFVAGGGSRDTFR 558
Db   523 ASPNAATTSGYTPLHLSAREGHEDVAAFLLDHGASLSITTKKGFTPLHVAA--------- 573

Qy   559 VVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENIL----GSPNAKFLQHIQSANF 614
Db   574 ----------------------KYGKLEVANLLLQKSASPDA--------AGKS 597

Qy   615 GFSPARRGIVSSNHNVMKDILNFVGDSLHLPSBRGYNAMQVAALFGDKEAVKMLAKSAKP 674
Db   598 GLTPLHVAAHYDNQKVALLLLD-QGASPHAAAKNGYTPLHIAAKKNQMDIATTLLEYG-- 654

Qy   675 SDLNFKTSATPTPLNLACLRGDNEVVRGLVGQH-GIDINQRMGSDKNTVLHYAISKGDSF 733
Db   655 ADANAVTRQGIASVHLAAQEGHVDMVSLLGRNANVNLSNKSGL---TPLHLA-AQEDRV 710

Qy   734 LVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVNRLDDNGRSVLSSA 793
Db   711 NTAEVLVNQGAHVDAQTKMGYTPLHVGCHYGNIKIVNPLLQHSAKVNAKTKNGYTPLHQA 770

Qy   794 IVPGRKEKGVLGIVNKLLDRGADIN 818
Db   771 -----AQQGHTHIINVLLQNNASPN 790

RESULT 11
```

```
Qy   231 -------ADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKALGEVENDEGDLGANAYNV 281
Db   233 VATLLLNRAAAVDFT----ARNDITPLH-VASKRGNANMVKLL------LDRGAK-IDA 279

Qy   282 LDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFQSPLHIITQKPDCSVLDIEE 341
Db   280 KTRDGLTPLHCGARSGHEQVVEMLLDRSAPILSKTKN--GLSPLHMATQGDHLNCVQL-- 335

Qy   342 FTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDPDYFGNTIAH 401
Db   336 LLQHNVPVDDVTNDYLTALHVAAHCGHYKVAKVLLD--KKASPNAKALN------GFTPLH 388

Qy   402 YAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVNSQD 461
Db   389 IACK-KNR-IRVMELLLKHGASIQAVTESGLTPIHVAAFMGHVNIVSQLMHHGASPNTTN 446

Qy   462 VNGDTPLHIAVEBGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIVGIKAL--GS 519
Db   447 VRGETALHMAARSGQAE-VVRYLVQDGAQVEAKAKDDQTPLHISARLGKADIVQQLLQQG 505

Qy   520 AKPNI-----------------KGEDTVA------KSLLMEDYKGFTPLHFVAGGGSRDTFR 558
Db   506 ASPNAATTSGYTPLHLAAREGHEDVAAFLLDHGASLSITTKKGFTPLHVAAKYGKLEV-- 563

Qy   559 VVRKNYEKCHDLATIRAALMQDRSGGELVNLGDFESENILGSPNAKFLQHIQSANFGFSP 618
Db   564 -------------ASLLLQKS---------------ASPDA------AGKSGLTP 584

Qy   619 ARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAKSAKPSDLN 678
Db   585 LHVAAHYDNQKVALLLLD-QGASPHAAAKNGYTPLHIAAKKNQMDIATSLLEYG--ADAN 641

Qy   679 FKTSATPTPLNLACLRGDNEVVRGLVGQH-GIDINQRMGSDKNTVLHYAISKGDSFLVQK 737
Db   642 AVTRQGIASVHLAAQEGHVDMVSLLLSRNANVNLSNKSGL---TPLHLA-AQEDRVNVAE 697

Qy   738 ILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVNRLDDNGRSVLSSAIVPG 797
Db   698 VLVNQGAHVDAQTKMGYTPLHVGCHYGNIKIVNPLLQHSAKVNAKTKNGYTALHQA---- 753

Qy   798 RKEKGVLGIVNKLLDRGADIN 818
Db   754 -AQQGHTHIINVLLQNNASPN 773

RESULT 10

A55575
ankyrin 3, long splice form - human
N;Alternate names: ankyrin G
C;Species: Homo sapiens (man)
C;Date: 03-Mar-1995 #sequence_revision 03-Mar-1995 #text_change 09-Jul-2004
C;Accession: A55575
R;Kordeli, E.; Lambert, S.; Bennett, V.
J. Biol. Chem. 270, 2352-2359, 1995
A;Title: Ankyrin-G. A new ankyrin gene with neural-specific isoforms localized at the ax
A;Reference number: A55575; MUID:95138209; PMID:7836469
A;Accession: A55575
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-4377 <KOR>
A;Cross-references: UNIPROT:Q12955; GB:U13616; NID:g608024; PIDN:AAA64834.1; PID:g6080025
C;Genetics:
A;Gene: GDB:ANK3
A;Cross-references: GDB:424503; OMIM:600465
A;Map position: 10q21-10q21
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
C;Keywords: alternative splicing; peripheral membrane protein
F;73-105/Domain: ankyrin repeat homology <AN01>
F;106-138/Domain: ankyrin repeat homology <AN02>
F;139-171/Domain: ankyrin repeat homology <AN03>
F;172-200/Domain: ankyrin repeat homology <AN04>
F;201-233/Domain: ankyrin repeat homology <AN05>
```

S37431

ankyrin 2, neuronal long splice form - human
N;Alternate names: ankyrin B, 440K splice form; ankyrin-B; brain ankyrin; non-erythroid
N;Contains: ankyrin 2, short form
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: S37431; S37431; A39643; B39643; A40334; A49462; S14533; S14569
R;Chan, W.
submitted to the EMBL Data Library, September 1993
A;Reference number: S37431
A;Accession: S37431
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-3924 <CHA>
A;Cross-references: UNIPROT:Q01484; EMBL:Z26634; NID:g406287; PIDN:CAA81387.1; PID:g4062
R;Otto, E.; Kunimoto, M.; McLaughlin, T.; Bennett, V.
J. Cell Biol. 114, 241-253, 1991
A;Title: Isolation and characterization of cDNAs encoding human brain ankyrins reveal a
A;Reference number: A39643; MUID:91302466; PMID:1830053
A;Accession: A39643
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2077 <OT1>
A;Cross-references: GB:X56957
A;Accession: B39643
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1443,3585-3924 <OTT>
A;Cross-references: EMBL:X56958
R;Tse, W.T.; Menninger, J.C.; Yang-Feng, T.L.; Francke, U.; Sahr, K.E.; Lux, S.B.; Ward,
Genomics 10, 858-866, 1991
A;Title: Isolation and chromosomal localization of a novel nonerythroid ankyrin gene.
A;Reference number: A49462; MUID:94075409; PMID:8253844
A;Accession: A40334; MUID:92009921; PMID:1833308
A;Accession: A40334
A;Molecule type: DNA
A;Residues: 463-474,'PE',477-495 <TSE>
A;Cross-references: GB:M37123; NID:g178647; PIDN:AAA62828.1; PID:g178648
R;Chan, W.; Kordeli, E.; Bennett, V.
J. Cell Biol. 123, 1463-1473, 1993
A;Title: 440-kD ankyrlin: structure of the major developmentally regulated domain and se
A;Reference number: A49462; MUID:94075409; PMID:8253844
A;Accession: A49462
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-3924 <RES>
A;Cross-references: EMBL:Z26634; NID:g406287; PIDN:CAA81387.1; PID:g406288
C;Genetics:
A;Gene: GDB:ANK2
A;Cross-references: GDB:127607; OMIM:106410
A;Map position: 4q25-4q27
C;Superfamily: ankyrin; ankyrin repeat homology
C;Keywords: alternative splicing
F;2-3924/Product: ankyrin 2, long form #status predicted <MAT>
F;63-95/Domain: ankyrin 2, short form #status predicted <MA2>
F;96-128/Domain: ankyrin repeat homology <AN01>
F;129-161/Domain: ankyrin repeat homology <AN02>
F;162-190/Domain: ankyrin repeat homology <AN03>
F;191-223/Domain: ankyrin repeat homology <AN04>
F;232-264/Domain: ankyrin repeat homology <AN05>
F;265-297/Domain: ankyrin repeat homology <AN06>
F;298-330/Domain: ankyrin repeat homology <AN07>
F;331-363/Domain: ankyrin repeat homology <AN08>
F;364-396/Domain: ankyrin repeat homology <AN09>
F;397-429/Domain: ankyrin repeat homology <AN10>
F;430-462/Domain: ankyrin repeat homology <AN11>
F;463-495/Domain: ankyrin repeat homology <AN12>
F;496-528/Domain: ankyrin repeat homology <AN13>
F;529-561/Domain: ankyrin repeat homology <AN14>
F;562-594/Domain: ankyrin repeat homology <AN15>
F;595-627/Domain: ankyrin repeat homology <AN16>
F;628-660/Domain: ankyrin repeat homology <AN17>
F;661-693/Domain: ankyrin repeat homology <AN18>
F;694-726/Domain: ankyrin repeat homology <AN19>
F;727-759/Domain: ankyrin repeat homology <AN20>
F;760-792/Domain: ankyrin repeat homology <AN21>
F;793-825/Domain: ankyrin repeat homology <AN22>
F;793-825/Domain: ankyrin repeat homology <AN23>

Query Match   8.5%; Score 368.5; DB 2; Length 3924;
Best Local Similarity   25.1%; Pred. No. 9.2e-13;
Matches 178; Conservative 87; Mismatches 262; Indels 183; Gaps 26;

Qy   184 TSGPVRDIADRIVASKGD----------LSEDQVEEILDIIFMNESEIAEGISNPLH------ 230
               ||||| : | : :.:   ::.: | |||||
Db   187 TKGKVRLPLHIAARKDDTKSAALLQNDHNADVQSKMVNRT--TESGFTPLHIAAHYG 244

Qy   231 ----------------------ADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKAL----GEVENDEGDL 274
                                  :|:|: | |.|:: |::   :::
Db   245 NVNVATLLNRGAAVDFT--------ARNGITPLH-VASKRGNTNMVKLLLDRGGQIDAKTRD- 298

Qy   275 GANAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSPLHIITQKPDC 334
                | | ||: |.|| :|: : : |: | |: | :. |||| :
Db   299 ------------GLTPLHCAARSGHDQVELLERGAPLLARTKN--GLSPLHMAAQ------- 341

Qy   335 SVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPV------------------ILK 373
                  : | :: || ||.|||| ||| |
Db   342 ------------GDHVECVKHLLQHKAPVDDVTLDYLTALHVAAHCGHYRVTK 382

Qy   374 GVMDHVKQNSSEFQDLVNDPDYFGNTIAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRA 433
               :.: | :: ::| ||:: || : |: :::
Db   383 LLLDKRANP-------NARALNGFTPLHIACK--KNR--IKVMELLVKYGASIQAITESGLT 433

Qy   434 PIHVASSNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSV 493
               |||| | || :.| :| :| |||||| |:: : : |:
Db   434 PIHVAAFMGHLNIVLLLQNGASPDVTNIRGETALHMAARAGQVEVVRCLL-RNGALVDA 492

Qy   494 QNNDGVTPMLSAAKYGDIGVIKAL--GSAKPNI------------------KGEDTVAKSLL- 535
               |:.||.||::|||:| : ::| |:| || | :| ||
Db   493 RAREEQTPLHIASRLGKTBIVQLLLQHMAHPDAATTNGYTPLHISAREGQVDVASVLLEA 552

Qy   536 -----MEDYKGFTPLHFVAGGSGRDTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLG 590
                     :.|| |||||:   :::   :
Db   553 GAAHSLATKKGFTPLHVAAKYGSLDVAKL---------------LLQRRAAAD-- 590

Qy   591 DFESENILGSPNAKFLQHIQSANFGFSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGY 650
               :
Db   591 ------------SAGKNGLTPLHVAAHYDNQKVALLLE-KGASPHATAKNGY 630

Qy   651 NAMQVAALFGDKEAVKMLAKSAKPSDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGID 710
               : |:| :: | |::: : : :| ||:
Db   631 TPLHIAAKKNQMQIASTLLNYG--AETNIVTKQGVTPLHLHASQEGHTDMVTLLLDK-GAN 687

Qy   711 INQRMGSDKN--TVLHYAISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKI 768
               |::|.   :   .|||:|:||| ||::|||:|:|::|| : | | ||||:||
Db   688 IHM--STKSGLTSLHLA-AQEDKVNVADILTKHGADQDAHTKLGYTPLLVACHYGNVKM 743

Qy   769 VSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADIN 818
               |:|||||||:||||| |||:|||||||| || || : :| |||||||
Db   744 VNFLLKQGANVNAKTKNGYTPLHQA-----AQQGHTHIINVLLQHGAKPN 788

RESULT 12

T42691
hypothetical protein DKFzp434D2328.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C;Accession: T42691
R;Bloecker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, November 1999
A;Reference number: Z22230
A;Accession: T42691
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-791 <AAA>
A;Cross-references: UNIPROT:Q9UFA4; EMBL:AL133087
A;Experimental source: adult testis; clone DKFzp434D2328

C;Genetics:
A;Note: DKFZp434D2328.1

Query Match   8.0%;   Score 345; DB 2; Length 791;
Best Local Similarity   23.0%; Pred. No. 1.7e-12;
Matches 174; Conservative 110; Mismatches 275; Indels 196; Gaps 31;

Qy   228 PLHADVDNNPVKGAKNVM-----------TLMHIVYACDVDPRIVKALGEVENDEGD 273
             ||| : | ::
Db   6 PLHAAASNGQINVVKHLLNLGVEIDEINVYGNTALHI--AC------YNGQDAVVNELID 57

Qy   274 LGANAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTDFIDT-ANFANQSPLHIITQK 331
             ||: ||||| : ||| ||
Db   58 YGANV-NQPNNNGFTPLHFAAASTHG--ALCLELLVNNGADVNIQSKDGKSPLHWTAVH 113

Qy   332 PDCSVLDIEEFT-SRNL-----DFGLVDGDGKNPLHHAVEHLPPVILKGVM--------- 376
                 : ||||| |||| || ::: |:
Db   114 -------GRFTRSQTLIQNGGEIDCVDKDGNTPLHVAARYGHELLINTLITSGADTAKC 165

Qy   377 ----------------------------------DHVKNSSEFQDLVNDPDYFGNT 398
                                               : ||: |:| ||||
Db   166 GIHSMPFLHLAALNAHSDCCRKLLSSGQKYSIVSLFSNEHVL-SAGFE--IDTPDKFGRT 222

Qy   399 IAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVASSNGKANAVSGLVSCGIDVN 458
             ||| ||:|| ::||| | | ||||| | |: | ||:| | |
Db   223 CLHAAAAGGNVE--CIKLLQSSGADFHKKDCGRTPLHYAAANCHFHCIETLVTTGANVN 280

Qy   459 SQDVNGDTPLHIA-----------VEGGSMET-------------VLAVLNQRGADV 491
             ||||||||| ||||| |||| |::| ||
Db   281 ETDDWGRTALHYAAASDMDRNKTILGNAHDNSEELERARELKEKEATLCLEFLLQNDANP 340

Qy   492 SVQNNDGVTPMLSAAKYG--------------DIGVIKA---LGSAKPNIKGEDT 529
             |:||||||| || || | | |||||| ||||
Db   341 SIRDKEGYNSIHYAAAYGHRQCLELLLERTNSGFEESDSGATKSPLHLAAYNGHHQALEV 400

Qy   530 VAKSLL---MEDYKGFTPLHFVAGGSRDTFRV-------------C 567
             || :|||||| || | | | |
Db   401 LLQSLVDLDIRDEKGRTALDLAAFKGHTECVEALINQGASIFVKDNVTKRTPLHASVING 460

Qy   568 HDLATIRAALMQDRSGGELVNLGDFESENIL----------GSPNAKFLQHIQSAN--- 615
             | : || | || || | | || ||| :|| |:|||
Db   461 HTLCL--RLLLEIADNPEAVDVKDAKGQTPLMLAVAYGHIDAVSLLLEKEANVDTVDILG 518

Qy   616 FSPARRGIVSSNHNVMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEAV-KMLAKSAKP 674
             ||||||||| || | | | | | |||| | || | || || ||| |: : |
Db   519 CTALHRGIMTGHEECVQMLLE-QEVSILCKDSRGRTPLHYAAARGHATWLSELLQMALSE 577

Qy   675 SDLNFKTSATPTPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDKNTVLHYAISKGDSFL 734
             :|| ||| |||| | || | | ||| || |||||||
Db   578 EDCCFADNQGVTPLHWACYNGNENCIEVLLEQKCP--RKFIGNPFTPLHCAIINDHGNC 634

Qy   735 VQKILAHTGVD---VNCENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRLDDNGRSVLS 791
             |||||| || |||||||||| ||| ||||| |||||| || | |||| |||
Db   635 ASLLLG--AIDSSIVSCRDDKGRTPLHAAAFADHVECLQLLLRHSAPVNAVDNSGKTALM 692

Qy   792 SAIVPGRKEKGVLGIVNKLLDRG-ADINL-DGDHN 824
             | : ||| || |:|| | ||| | |||
Db   693 MA------AENGQAGAVDILVNSAQADLTVKDKDLN 722

RESULT 13
S11527
alpha-latrotoxin precursor - black widow spider
C;Species: Latrodectus mactans tredecimguttatus (black widow spider)
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 16-Aug-2004
C;Accession: S11527; JN0313
R;Kiyatkin, N.I.; Dulubova, I.E.; Chekhovskaya, I.A.; Grishin, E.V.
FEBS Lett. 270, 127-131, 1990
A;Title: Cloning and structure of cDNA encoding alpha-latrotoxin from black widow spider
A;Reference number: S11527; MUID:91031994; PMID:1977615
A;Accession: S11527
A;Molecule type: mRNA
A;Residues: 1-1401 <KIY>

A;Cross-references: EMBL:X55009; NID:g9568; PIDN:CAA38753.1; PID:g9569
A;Note: it is uncertain whether Met-1, Met-8 or Met-18 is the initiator
R;Volkova, T.M.; Galkina, T.G.; Kudelin, A.B.; Grishin, E.V.
Bioorg. Khim. 17, 437-441, 1991
A;Title: Tryptic fragments structure of the black widow spider venom neurotoxin.
A;Reference number: JN0313; MUID:91362695; PMID:1888339
A;Note: English title and abstract; article in Russian
A;Accession: JN0313
A;Molecule type: protein
A;Residues: 21-33,'X',35-40,'X',43-50,82-90,'X',92-99,107-113;114-126;146-151;171-178;1
84-792;881-891;906-913;914,1015-1021;1062-1065,'X',1067-1075;1080-1084;1087-1096;1183-1
A;Note: the sequence of the amino terminal peptide has been corrected with tryptic pept
C;Superfamily: ankyrin repeat homology; EGF homology
C;Keywords: neurotoxin; venom
F;1-20/Domain: propeptide #status predicted <PRO>
F;21-1401/Product: alpha-latrotoxin #status experimental <MAT>
F;525-557/Domain: ankyrin repeat homology <AN7>
F;862-894/Domain: ankyrin repeat homology <AN3>
F;1005-1036/Domain: ankyrin repeat homology <AN21>

Query Match   8.0%;   Score 344; DB 2; Length 1401;
Best Local Similarity   22.4%; Pred. No. 4.9e-12;
Matches 138; Conservative 97; Mismatches 207; Indels 174; Gaps 24;

Qy   289 PLHHAAKNCTGDKLKLCMEKTKTDFIDTANFANQSP--------------LHIIT 329
             ||||| :   : :: : ::   :
Db   405 PVSVREKAC--PTLRMPVDQTRRNVLVFRKFDSSKPQLVGEITPYLSNFIDIRDLYNAA 462

Qy   330 QKPDCSVLDIEEFTSRNLDFGLVDGDGKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLV 389
             | | :|||| ||
Db   463 SNPD-SAVGFKEPTKLNY------DGAN--------------------IRATFDH---- 490

Qy   390 NDPDYFGNTIAHYAVKNKNADLTLFNM-LKASGADLNVRNVVGRAPIHVASSNGKANAVS 448
                   || | | |:: : : | : | ||||||| || |||| || |:|| | | |
Db   491 ------GRTVFHAAAKSGN-DKIMFGLTFLAKSTELNQPDKKGYTPIHVAADSGNAGIVN 543

Qy   449 GLVSCGIDVNS--------------------------QDVNGDTPLHIAVE 473
             || |||| || |||||||| |
Db   544 LLIQRGVSINSKTYHFLQTPLHLAAQRGFVTTFQRLMESPEININERDKDGFTPLHYAIR 603

Qy   474 GGSMETVL-AVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVIKA-LGSAKPNIKGEDTVA 531
             || | | || |||| || ||| | || | | |:: ||||||||||
Db   604 GG--ERILEAFLNQISIDVNAKSNTGLTPFHLAIIKNDWPVASTLLGSKKVDINAVDE-- 659

Qy   532 KSLLMEDYKGFTPLHFVAGGSRDTFRVVRKNYEKCHDLATIRAALMQDRSGGELVNLGD 591
             |:||| :| |||||| || || | :|:|
Db   660 -------NNITALHYAAILGYLETTK----------------------QLINLKE 685

Qy   592 FESENILGSPN------------------------AKFLQHIQSANF------GFSPARRGIVSSNHN 629
             ||||:|| | ||| |:||||| ||||||||| :||
Db   686 INA-NVVSSPGLLSALHYAILYKHDDVASFLMRSSNVVNLKALGGITPLHLAVIQGRKQ 744

Qy   630 VMKDILNFVGDSLHLPSERGYNAMQVAALFGDKEAVRMLAKSAKPSDLNFKTSATPTPLN 689
             ||| | || || | | : | || | : | | | || : |:|| |
Db   745 IL-SLMFDIGVNIEQKTDEKYTPLHLAAMSKYPELIQLLLDQG--SNFEAKTNSGATPLH 801

Qy   690 LACLRGDNEVVRGLVGQHGIDINQRMGSDKN--TVLHYAISKGDSFLVQKILAHTGVDVN 747
             || | | | ||||| || | | :|:| | | |:| || |||| || || |
Db   802 LATFGKGSQAALILLNN---EVNWR-DTDENGQMPIHGAAMTGLLDVAQAIISDATVVD 857

Qy   748 CENNLGQTPLHLAVEGGDPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIV 807
             |||||||||| || | |||: || || | | || : | | : | |
Db   858 IEDKNSDTPLNLAAQNSHIDVIKYFIDQGADINTRNKKGLAPLLAF-----SKKGNLDMV 912

Qy   808 NKLLDRGADINL-DGD 822
             | | | || | | |
Db   913 KYLFDKNANVYIADND 928

RESULT 14
H71274
probable ankyrin - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: H71274
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71274; MUID:98332770; PMID:965876
A;Accession: H71274
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-934 <COL>
A;Experimental source: strain Nichols
A;Cross-references: UNIPROT:O83807; GB:AE001254; GB:AE000520; NID:g3323148; PIDN:AAC6580
C;Genetics:
A;Gene: TP0835
C;Superfamily: syphilis spirochete probable ankyrin; ankyrin repeat homology
F;281-313/Domain: ankyrin repeat homology <AN22>

Query Match   7.9%;   Score 342; DB 1;   Length 934;
Best Local Similarity 26.5%;   Pred. No. 3.3e-12;
Matches 177; Conservative 85; Mismatches 241; Indels 166; Gaps 31;

Qy   226 SNPLHADVDNNPVKGA------KNVM------TLMHLYYACDVDPRIVKALGEVENDE 271
                :|| :|| | | | | :|:| ||| ||| | || ||
Db   226 SNPLHADVDNNPVKGA-AIKDDYGETPLHVAA------HIITQKPDCSVLDIEEFT---- 335

Qy   272 GDLGANAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTD----FIDTANFANQSP 324
               |||||| ||| |||||||||| || ||| |:|:| |||||| : |
Db   283 ATPLHEAVRAGQVDAAVLLLRSGADPNVRDASGNTCLHLVAPAFPRVRLVGAL--- 335

Qy   336 LDAGASV-AIKDDYGETPLHVAA---------RLGMDRAFVERLVGAGADISERNKKGETP 386
               ||||:| |||| ||||||| || : | || ||| | | |
Db   272 GDLGANAYNVLDSEGNLPLHHAAKNCTGDKLKLCMEKTKTD-------FIDTANFANQSP 324

Qy   325 L--------HIITQKPDCSVLDIEEFT------SRNL----DFGLVDGD-- 355
               | : |:| | :|:| || | | || :|||
Db   336 LDAGASV-AIKDDYGETPLHVAA---------RLGMDRAFVERLVGAGADISERNKKGETP 386

Qy   387 LVLTIDRDHRDLTAYFVSLGADIHAEDMRGETPLTKALARGLETVKIVVTDSNLYKQDVV 446
                ||| | : :||| ||| |||| | |||| | | | : |: ||| | || |
Db   325 L--------HIITQKPDCSVLDIEEFT-------SRNL----DFGLVDGD-- 355

Qy   356 GKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDP-----DYFGNTIAHYAVKNKNA 409
               |||||||||||| ||||||||||||| ||||| ||| | |||| | |:|
Db   387 LVLTIDRDHRDLTAYFVSLGADIHAEDMRGETPLTKALARGLETVKIVVTDSNLYKQDVV 446

Qy   447 GRDPLHVAVSR------RAHLDIVK------FLFREPKQMIARDTMGNTLLHYAVAND-- 492
               |:|||||| | |: |:| : ||| | | |:|||:|||
Db   356 GKNPLHHAVEHLPPVILKGVMDHVKNSSEFQDLVNDP-----DYFGNTIAHYAVKNKNA 409

Qy   410 DLTLFNMLKASGADLNVRNVVGRAPIHVA--SSNGKANAVSGLVSCGIDVNSQDVNGDTP 467
               |:| :||||||| | |:|| :||| | | | ||| |||| || |||| |
Db   447 GRDPLHVAVSR------RAHLDIVK------FLFREPKQMIARDTMGNTLLHYAVAND-- 492

Qy   493 DRAVGEFLMREGADIFSTNVHGVSPLKTALTTSGGRED-----WILTAANVHAQDTGGNTP 548
               | | | ||| |||| :|| |||| || | || | || | | | |
Db   410 DLTLFNMLKASGADLNVRNVVGRAPIHVA--SSNGKANAVSGLVSCGIDVNSQDVNGDTP 467

Qy   468 LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVI--------- 514
               ||:|| | | |: |::|||||| :| ||| |:| | ||| |:
Db   493 DRAVGEFLMREGADIFSTNVHGVSPLKTALTTSGGRED-----WILTAANVHAQDTGGNTP 548

Qy   549 LHLACEWKLTQAINGIL-RKGAEIEARNLNQETPLFSAVKSDAAEVISILLHPQAGNPAL 607
               ||:| |:|| |: ||| :|| || |:||| |||| | | |: || || | |
Db   468 LHIAVEGGSMETVLAVLNQRGADVSVQNNDGVTPMLSAAKYGDIGVI--------- 514

Qy   515 ------------KALGSAKPNIKGEDTVAKSLL-MEDYKGFTPLHFVAGGGSRD 555
                           ||||||||||||||| |||| :|||| |||||| ||| |||
Db   549 LHLACEWKLTQAINGIL-RKGAEIEARNLNQETPLFSAVKSDAAEVISILLHPQAGNPAL 607

Qy   608 VDARDAVGNTVLHACVRWSALRSADVLIREADARHVSLLNARNLSGKPPLHLAARAGNVD 667
               | | ||||| ||| | | | | |:: | || :|||| |||| ||| |||||
Db   515 ------------KALGSAKPNIKGEDTVAKSLL-MEDYKGFTPLHFVAGGGSRD 555

Qy   556 TFRVVRKNYEKCHDLATIRAAL-MQDRSG---GELVNLGDFESENILGSPNAKFLQHIQ 610
               | :| | | | | | ::||| |||: || |||||||||||||||| |||||
Db   608 VDARDAVGNTVLHACVRWSALRSADVLIREADARHVSLLNARNLSGKPPLHLAARAGNVD 667

Qy   668 FIRL--------LLSHRVALHMGDETGKSALTDAVLADQEESVHMLLSAGA---NPVQ 714
               |:| |||:|||| | | ||| | || |||:|:| |||||| |||
Db   556 TFRVVRKNYEKCHDLATIRAAL-MQDRSG---GELVNLGDFESENILGSPNAKFLQHIQ 610

Qy   611 SANFGFSPARRGIVSSNHNVMKDIINFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAK 670
               ||: |||| ||||||||| ||||:||||||:||||||||||| |||||| ||| || |
Db   668 FIRL--------LLSHRVALHMGDETGKSALTDAVLADQEESVHMLLSAGA---NPVQ 714

Qy   715 QDMYGRTPLHEAVLCNSQSVIA-ALRAAGNPFARDSYGTTPLSLALLKGDTFVGAVVGK 773
               ||: | ||| ||| | | |: || |||| || | ||||||: || || | | | |
Db   611 SANFGFSPARRGIVSSNHNVMKDIINFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAK 670

Qy   671 SAKPSDLNFKTSATPTPLNLACLRGDNEVVRG---LVGQHGIDINQRMGSDKNTVLHYA 726
               ||||||||||||||||||||||||||||| || | |||||| ||||| | ||||||
Db   715 QDMYGRTPLHEAVLCNSQSVIA-ALRAAGNPFARDSYGTTPLSLALLKGDTFVGAVVGK 773

Qy   774 ------DPLLANSDGQTPLHLAVM-ENVPVQTFRLLLAKGYPIDKRDRMGS---SALVLA 823
                     || | || ||||||||| |:||| | || ||| ||| | | || | |
Db   671 SAKPSDLNFKTSATPTPLNLACLRGDNEVVRG---LVGQHGIDINQRMGSDKNTVLHYA 726

Qy   727 ISKGDSFLVQKILAHTGVDVNCENNLGQTPLHLAVEGGDPKIVSSLLK-AGAVVNRLDD 784
               |||||||||| |:|| || ||| || |||||||||||| ||| ||| || ||| |
Db   774 ------DPLLANSDGQTPLHLAVM-ENVPVQTFRLLLAKGYPIDKRDRMGS---SALVLA 823

Qy   785 NGRSVLSSA 793
               : ||:| ||
Db   824 IKKDRSDJCHELLA-LGADLFIANNVGESPALLVLSKN-----TSILKTLVGPAVNKTDS 877

Qy   878 AGESILHYA 886
Db   785 NGRSVLSSA 793

Db   878 AGESILHYA 886

RESULT 15

T13940
ankyrin - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T13940
R;Dubreuil, R.R.; Yu, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 10285-10289, 1994
A;Title: Ankyrin and beta-spectrin accumulate independently of alpha-spectrin in Drosopt
A;Reference number: Z17820; MUID:95024098; PMID:7937942
A;Accession: T13940
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1549 <DUB>
A;Cross-references: UNIPROT:Q24241; EMBL:L35601; NID:g557083; PID:g557084; PIDN:AAC37208
C;Genetics:
A;Cross-references: FlyBase:FBgn0011747

Query Match   7.9%;   Score 339.5; DB 2;   Length 1549;
Best Local Similarity 22.9%;   Pred. No. 1.1e-11;
Matches 164; Conservative 100; Mismatches 249; Indels 203; Gaps 27;

Qy   227 NPLHADVDNNPVKGAKNVMTLMHLVYACDVDPRIVKAL-GEVENDEGDLGANAYNVLDS 284
                :|| | ::||| : : | | || |||| | ||:|: | : |||||| |
Db   22 NGMALDNKNGIIKQNDATISFLRAARSGDI-KKVMDFLDCGEI-SDINSCNANGLNA--- 76

Qy   285 EGNLPLHHAAKNCTGDKLKLCMEKTKTDF-IDTANFANQSPLHIITQKPDCSVLDIEEFT 343
                |||||||||| | | |||| ||||||| : |||| | | |||:| | || | | |
Db   77 -----LHLAAKDGYVD---ICCELLRRGIKIDNATKKGNTALHIASLAGQHDVINQLILY 128

Qy   344 SRNLDFGLVDG------------------------DGKNPLHHAVEHLPPVILKGVMDHVKNS 382
                | | :| || | | |||||||||| |||||| | |::
Db   129 NANVNVQSLNGFTPLYMAAQENHDNCCRTLLANGANPSLSTEDGFTPLAVAMQQGHDK--- 186

Qy   383 SEFQDLVNDPDYFGNT---IAHYAVKNKNADLTLFNMLKASGADLNVRNVVGRAPIHVAS 439
               : |||:|:| ||: : |||| ||:|| || | |:||||| || ||: || || |
Db   187 --IVAVLLENDVRGKVRLPALHIAAK--KN-DVNAAKLLQHDPNADIVSKSGFTPLHIAA 242

Qy   440 SNGKANAVSGLVSCGIDVNSQDVNGDTPLHIAVEGGSMETVLAVLNQRGADVSVQNNDGV 499
               | | | ||| |||| || |||| ||||| || | : | | |::| |: | || |
Db   243 HYGNVDIATLLNNKADVNYVAKHNITPLHVACKKGKL-SLCTLLLCRGAKIDAATRDGL 301

Qy   500 TPMLSAAKYGDIGVIK------------------------ALGSAKPNIKGEDTVAKSLLME- 537
               ||| |||| ||: |:| ||||||||||||||| |||
Db   302 TPLHCASRGSHVEVIKHLLQQNAPILTKTKNGLSALHMAAQGEHDEAAHLLLDNKAPVDE 361

Qy   538 ------------------DYK------------------GFTPLHFVAGGGSRDTFRVVRKN 563
                                 |:| |||||| ||| |||| :| | |
Db   302 TPLHCASRGSHVEVIKHLLQQNAPILTKTKNGLSALHMAAQGEHDEAAHLLLDNKAPVDE 361

Qy   362 VTVDYLTALHVAAHCGHVKVAKLLLDYKANPNARALNGFTPLHIAC------------KKN 410
               |||| : |||: | |||| |||||| |||| | :|||||||||| |||
Db   538 ------------------DYK------------------GFTPLHFVAGGGSRDTFRVVRKN 563

Qy   564 YEKCHDL-----ATIRAALMQDRSGGELVNLGDFESENILGSPN--AKFLQHIQSANFGF 616
                    ||| ||| |:|||: ||||||| |||||||||||| || ||||: ::|
Db   362 VTVDYLTALHVAAHCGHVKVAKLLLDYKANPNARALNGFTPLHIAC------------KKN 410

Qy   411 RIKMVELLIKHGANIGAT----TESGLTPLHVASF-----MGCINIVIYLLQHEASAD--- 459
               | |:|| | |||| ||| || ||||||||| :|| || : || |||
Db   564 YEKCHDL-----ATIRAALMQDRSGGELVNLGDFESENILGSPN--AKFLQHIQSANFGF 616

Qy   617 SPARRGIVSSNHNVMKDIINFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAKSAKPSD 676
               |||| |||||||| |||: ||||||:||||||||||||| ||||| || |||||||| |
Db   411 RIKMVELLIKHGANIGAT----TESGLTPLHVASF-----MGCINIVIYLLQHEASAD--- 459

Qy   460 ------------------LPTIRGETPLHLAARANQDIIRILLRSAKVDA 492
                                 |||: ||||||||| | :| |:|| ::| |
Db   617 SPARRGIVSSNHNVMKDIINFVGDSLHLPSERGYNAMQVAALFGDKEAVKMLAKSAKPSD 676

Qy   677 LNFKTSATPTPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDKNTVLHYAISKGDSFLVQ 736
               ||||||||| |||||| | |:| | ||| |||||||| || ||||||:| | |:|:|
Db   460 ------------------LPTIRGETPLHLAARANQDIIRILLRSAKVDA 492

Qy   493 I---VREGQTPLHVASRLGNINIIM-LLLQHGAEINAQ-SNDKYSALHIAAKEGQENIVQ 547
                   : | |||||||| | :|:|:: ||| || ||| | ||:| || ||| | |:|: |
Db   677 LNFKTSATPTPLNLACLRGDNEVVRGLVGQHGIDINQRMGSDKNTVLHYAISKGDSFLVQ 736

Qy   737 KILAHTGVDVNCENNLGQTPLHLAVEGG 764
               |:|| || ||| | | | ||||||| ||
Db   493 I---VREGQTPLHVASRLGNINIIM-LLLQHGAEINAQ-SNDKYSALHIAAKEGQENIVQ 547

Qy   548 -VLLENGAENNAVTKKGFTPLHLACKYGKQNVVQILLQNGASIDFQGKNDVTPLHVATHY 606
                : | | |:||||::|||||| | || ||:| ||| |||||| ||| || ||||| :|
Db   737 KILAHTGVDVNCENNLGQTPLHLAVEGG------SALVLA 764

Qy   765 -DPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADINL 819
                || |||||| | |:| ||| ||||||||||:| |::||||| | |||||||||||
Db   548 -VLLENGAENNAVTKKGFTPLHLACKYGKQNVVQILLQNGASIDFQGKNDVTPLHVATHY 606

Qy   607 NNPSIVELLLKNGSSPNLCARNGQCAIHIAC-----KGNYLEIAMQLLQHGADVNI 657
Db   765 -DPKIVSSLLKAGAVVNRLDDNGRSVLSSAIVPGRKEKGVLGIVNKLLDRGADINL 819

Db   607 NNPSIVELLLKNGSSPNLCARNGQCAIHIAC-----KGNYLEIAMQLLQHGADVNI 657